Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
PAUL M. FITTS, *University of Michigan*
FREDERICK C. FRICK, *Massachusetts Institute of Technology*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
CLARENCE H. GRAHAM, *Columbia University*
HAROLD W. HAKE, *University of Illinois*
LLOYD G. HUMPHREYS, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

A Stimulus-Trace Hypothesis for Statistical Learning Theory: R. S. WITTE 273
The Relation Between Mean Reward and Mean Reinforcement: A. M. LEVENTHAL, R. F. MORRELL, E. F. MORGAN, JR., AND C. C. PERKINS, JR. 284
A Study of Semantic Generalization Through the Use of Established Conceptual Mediations: R. J. KAPLAN 288
Comparison of Predictions and Estimates in a Probability Learning Situation: E. D. NEIMARK AND E. H. SHUFORD 294
Some Conditions Governing the Use of the Cue-Producing Response as an Explanatory Device: J. DE RIVERA 299
Changes in Performance as a Function of Shifts in the Magnitude of Reinforcement: G. COLLIER AND M. H. MARX 305
Growth of a Motor Skill as a Function of Distribution of Practice: J. M. DIGMAN 310
Response Decrement, Induced by Stimulus Change, as a Function of Amount of Training: D. BINDRA AND J. F. SEELY 317
Stimulus Generalization as a Function of Level of Motivation: D. R. THOMAS AND R. A. KING. 323
The Relation Between Physiological Measures of Emotionality and Performance in Eyelid Conditioning: W. N. RUNQUIST AND L. E. ROSS 329
Set to Respond and the Effect of Interrupting Signals upon Tracking Performance: S. GRIEW.. 333
Effects of Inversion of the Visual Field on Human Motions: W. RHULE AND K. U. SMITH.... 338
The Sensitivity of the Eye to Two Kinds of Continuous Transformation of a Shadow-Pattern: K. VON FIEANDT AND J. J. GIBSON 344
Replication Report: The Relationship of Manifest Anxiety and Electric Shock to Eyelid Conditioning: D. F. CALDWELL AND R. L. CROMWELL 348
Supplementary Report: Discrimination Learning with Probabilistic Reinforcement Schedules: R. C. ATKINSON, W. H. BOGARTZ, AND R. N. TURNER 349

American Psychological Association

Vol. 57, No. 5 May 1959

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $8.00, or $16.00 annually. Single copies are $1.50. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, 53, 1–2).

Address all articles submitted for publication to the Editor, Arthur W. Melton, Department of Psychology, University of Michigan, Ann Arbor, Michigan. **Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*.** When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1959 the publication cost is $30 per printed page. Authors of priority publications receive no free offprints.

To save printing costs, supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

ARTHUR C. HOFFMAN, Managing Editor
HELEN ORR, Promotion Manager
BEATRICE B. HAMLIN, Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1959 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 57, NO. 5 MAY, 1959

A STIMULUS-TRACE HYPOTHESIS FOR STATISTICAL LEARNING THEORY [1]

ROBERT S. WITTE [2]

Stanford University

A recent extension of the verbal conditioning paradigm to sequentially dependent reinforcements has exposed a limitation of statistical learning theory. This study attempts to remove that limitation by testing an extension of the theory.

Statistical learning theory yields several deductions, two of which will be considered here, for the asymptotic behavior of *S*s in verbal conditioning situations. According to the theory, *S*s should predict a given reinforcing event about as frequently as the event occurs. A number of experiments (e.g., Estes & Straughan, 1954; Goodnow & Postman, 1955) lend support to the "probability matching" deduction. The theory also predicts a slight sequential dependency between *S*s' predictions on Trial n and the reinforcing event on Trial $n - 1$. Tests of this deduction (Anderson & Grant, 1957; Bottenberg, 1956; Estes & Straughan, 1954) have found larger sequential dependencies than forecast by the theory.

Following the lead of Hake and Hyman (1953), Lawrence employed a verbal conditioning design in which successive reinforcements were sequentially dependent.[3] As required by the theory, his *S*s predicted a given reinforcing event about as frequently as the event occurred, but they also structured their predictions in a manner reflecting the sequential dependencies between reinforcements. For example, in one of Lawrence's groups each of two reinforcing events, E_1 and E_2, occurred 50% of the time. In addition, E_1 followed E_1 90% of the time, while E_2 followed E_1 10% of the time. The *S*s working under this schedule not only predicted E_1 50% of the time, but they also predicted E_1 approximately 90% of the time on trials following E_1 occurrences. Except for the slight sequential dependency mentioned above, the probability of E_1 predictions on trials following E_1 occurrences should approach .50. This was clearly not the case in the study

[1] This study is part of a thesis submitted to the Department of Psychology of Stanford University in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to thank Douglas H. Lawrence, who acted as sponsor, and William K. Estes, Richard C. Atkinson, and Quinn McNemar for their valuable suggestions and criticisms.

[2] Now at Cornell College, Mount Vernon, Iowa.

[3] Unpublished study by D. H. Lawrence. Personal communication. July, 1956.

by Lawrence or in more recently reported studies (Anderson, 1956; Engler, 1958).

Inasmuch as the above experimenters used relatively massed trials, the stimulus confronting S on any trial may have been a complex one containing traces of previous responses and reinforcements. As suggested by Estes and Straughan (1954), assume that the stimulus includes a trace of the reinforcing event from the previous trial. Utilizing this trace assumption, Burke and Estes (1957) and the writer (1958) derived expressions for conditional and marginal response probabilities which closely approximate the results obtained when sequentially dependent reinforcements were used.

The extended model assumes that no overlap occurs between the stimulus elements which are present on trials following E_1 and E_2 events. If trial spacing is distributed, this assumption becomes untenable and the model must be modified. A reinterpretation of Estes' (1955) theoretical formulation for spontaneous regression permits the prediction of the effect of trial spacing on conditional response probability.

Three mutually exclusive sets of stimulus elements, S^*_e, S^*_1, and S^*_2, can be defined so that each set contains two mutually exclusive and exhaustive subsets, S_e and S'_e; S_1 and S'_1; and S_2 and S'_2, respectively. Let S_e represent the set of all available elements on a given trial, except for those available elements associated with the reinforcing event of the previous trial. Typically, S_e is identified with postural and situational cues which are present at the start of any trial. Let S_1 represent the set of available elements which supplements S_e on trials following an E_1 occurrence, and let S_2 represent the set of available elements which supplements S_e on trials following an E_2 occurrence. In the present model, S_1 and S_2 are identified with those available elements contributed by the hypothesized trace of the previous reinforcement.

For the purpose of this paper, one can assume that S_e contains a constant number of elements N_e at the start of each trial, regardless of the size of the intertrial interval. Now, however, assume that the N_1 elements in S_1 and the N_2 elements in S_2 fluctuate randomly between S_1 and S'_1 and between S_2 and S'_2, respectively, during the intertrial interval. During a unit of time Δt following an E_1 occurrence, there exists a probability j_1 that any element in S_1, the available set, will escape into S'_1, the unavailable set. During the same interval there exists a probability j'_1 that any element in S'_1 will go into S_1. Then the general solution for $F_1(t)$, the probability that any element is in S_1 after the (t)th interval, is

$$F_1(t) = \frac{j'_1}{j_1 + j'_1} - \left[\frac{j'_1}{j_1 + j'_1} - F_1(0)\right] \times (1 - j_1 - j'_1)^t \quad [1]$$

where $F_1(0)$ represents the value of $F_1(t)$ when E_1 occurred.

Since S_1 is identified with a stimulus trace, it is reasonable to assume that j'_1 equals zero. Thus, Equation 1 reduces to

$$F_1(t) = F_1(0)(1 - j_1)^t \quad [2]$$

Defining $N_1(t)$ as the number of elements in S_1 after the (t)th interval and substituting $N_1(t)$ for $F_1(t)$ in Equation 2, one obtains

$$N_1(t) = N_1(1 - j_1)^t \quad [3]$$

Then $p_{11}(n,t)$, the probability of an E_1 response after Trial n, given an E_1 occurrence of Trial n and an intertrial interval of size t, can be expressed as

$$p_{11}(n,t) = \frac{1}{N_e + N_1(t)} \times \{N_e[(1-\theta)F_e(n-1)+\theta] + N_1(t)[(1-\theta\Pi_{11})F_1(n-1)+\theta\Pi_{11}]\} \quad [4]$$

where θ is a sampling parameter; Π_{11} is the probability of an E_1 event following an E_1 event; $F_e(n-1)$ is the probability that the ith element in S_e is conditioned to the E_1 response class after Trial $n-1$; and $F_1(n-1)$ is the probability that the ith element in S_1 is conditioned to the E_1 response class after Trial $n-1$, given an E_1 occurrence on Trial $n-1$.

When rewritten in terms of asymptotic behavior, Equation 4 takes the form

$$p_{11}(\infty,t) = \frac{1}{N_e + N_1(t)} \times \{N_e[\Pi + \theta(1-\Pi)] + N_1(t)[\Pi_{11} + \theta\Pi_{11}(1-\Pi_{11})]\} \quad [5]$$

where Π is the probability of an E_1 event, and the terms in the square brackets now represent the asymptotic values of $F_e(n-1)$ and $F_1(n-1)$, plus small "correction" terms required by the consistent incremental effect of the previous E_1 occurrence. According to Equation 5, when $N_1(t)$ is comparatively large, conditional response probability approaches Π_{11} plus a term smaller than θ in magnitude, and when $N_1(t)$ is comparatively small, it approaches Π plus a small term less than θ in magnitude.

An expression analogous to that for $p_{11}(\infty,t)$ can be obtained for $p_{21}(\infty,t)$, the asymptotic probability of an E_1 response after Trial n, given an E_2 occurrence on Trial n and an intertrial interval of size t:

$$p_{21}(\infty,t) = \frac{1}{N_e + N_2(t)} \times \{N_e[\Pi - \theta\Pi] + N_2(t)[\Pi_{21} - \theta\Pi_{21}(1-\Pi_{21})]\} \quad [6]$$

where Π_{21} is the probability of an E_1 event following an E_2 event. Except for a change in notation and in the correction terms, Equation 6 has the same properties as Equation 5.

The remaining part of this paper describes an attempt to test the trace-fluctuation model by varying the intertrial interval and noting its effect on conditional response probability.

Method

Design.—Practical considerations dictated the use of rats as *Ss*; *E* anticipated much difficulty in motivating college sophomores to participate in that phase of the experiment requiring highly distributed trials. Fortunately, the identifications between theory and data for the verbal conditioning situation can be transposed readily to the T-maze situation with rats. More specifically, a food reward in the left or right goal box was identified with an E_1 or E_2 event, while a left or right turn at the choice point was identified with an E_1 or E_2 response.

By definition, the intertrial interval began when *S* was lifted from the goal box, and it ended when *S* reached the choice point on the following trial. Since a pilot study indicated that the time required by most *Ss* to travel from the starting box to the choice point stabilized at about 2 sec., the size of the intertrial interval could be predetermined within a small margin of error. Each *S* was assigned permanently to one of four intertrial intervals: 8 sec., 30 sec., 3 min., or 1 hr.

All *Ss* were rewarded equally often on both sides of the maze. One-half of the experimental *Ss* worked under a "repetition" reward schedule in which E_1 followed E_1 90% of the time, while E_1 followed E_2 10% of the time. To compensate for possible spontaneous alternation tendencies (Dember & Fowler, 1958), the other one-half of the experimental *Ss* worked under an "alternation" reward schedule in which E_1 followed E_1 10% of the time, while E_1 followed E_2 90% of the time. As an additional check on alternation tendencies, a control group worked at the 8-sec. intertrial interval under a reward schedule in which E_1 followed E_1 and E_2 equally often. Henceforth, the various groups of experimental *Ss* will be designated as 8-R, 8-A, etc., where the number refers to the size of the intertrial interval in seconds and the letter refers to the reward schedule.

Apparatus.—The apparatus, a simple T maze, consisted of a gray stem, 28 in. long, and two black alleyways, each 36 in. long. All sections of the maze were 4 in. wide and 4 in. high. The stem contained two guillotine gates, one at the starting box and the other at the choice point. In each alleyway, at a distance of 21 in. from the choice point, a hinged door separated the alleyway from the goal box. These doors could be locked by

concealed levers. The levers were operated mechanically from *E*'s stand at the base of the maze.

Approximately 3 ft. above the maze, a set of eight 100-w. bulbs were located as follows: one bulb above the starting gate, another above the choice point, and the remaining were arranged symmetrically above the two branch alleyways. No other light source was present in the experimental room.

The intertrial interval for the 8- and 30-sec. groups was timed to the nearest second with an automatic recycling timer. To facilitate the accuracy of this measurement, button switches controlling the timer were located adjacent to the goal boxes. The interval for the 3-min. group was timed with a stop watch to the nearest 5 sec., and that for the 1-hr. group was timed with a wall clock to the nearest minute.

Procedure.—During the first six days of a 12-day training period, *Ss* were gentled and introduced to a 22½-hr. food privation schedule. During the last six days of the training period, *Ss* explored each of the two alleyways of the T maze on alternate days, learning to run through the goal box door and locate a wet mash reward in the foodwell.

After the training period, trials in the T maze were begun. An additional 12 days elapsed before *Ss*' running times permitted the desired five trials per day. On each trial, after *S* had been placed in the starting box, the gates at the starting box exit and at the choice point were raised. When *S*'s entire body penetrated a given alleyway, these gates were lowered, and one of the goal box doors was unlocked according to a predetermined schedule. If *S* ran to the side containing the locked door, he was forced to return to the choice point and proceed to the other side. After *S* had eaten the mash reward, weighing approximately .3 gm., he was lifted from the goal box, and the goal box door was relocked. The *Ss* in the 8-sec. group were returned immediately to the starting box and run for the next trial. The *Ss* in the remaining groups were placed in individual detention cages for prescribed periods of time before receiving the next trial.

To further motivate a few *Ss* who occasionally froze in the maze, all *Ss* who refused to run on any trial lost their 1½-hr. free-feeding privileges for that day. This technique was required only during the first three days of the experiment. Ordinarily, food deprivation at running time ranged between 18 and 22 hr. All groups participated in a counterbalanced design yielding an average of 20-hr. food deprivation at running time.

The *Ss* in the control group received a total of 200 trials. In an attempt to give all experimental *Ss* an ample series of trials in which to discriminate the sequential dependencies between reinforcements, the experimental *Ss* at the 8- and 30-sec. intervals received 100 trials more than did the control *Ss*, while *Ss* at the 3-min. and 1-hr. intervals received 200 trials more than did the control *Ss*.

The reward schedule was predetermined separately for each *S* over blocks of 100 trials. These schedules were constructed from tables of random numbers with the restriction that the reward sequence reflected within a margin of 2% the conditional probabilities specified in the experimental design.

Subjects.—The *Ss* were 171 albino rats between 90 and 140 days of age at the start of the experiment. Each of the groups in the study received approximately proportionate allotments of the 120 males and 51 females. Twelve fatalities occurred during the experiment. This high mortality rate was primarily caused by respiratory complications, first detected during the training period and subsequently treated with sulfathiazole. Table 3 shows the number of *Ss* in each group at the end of the experiment.

Results

Unless noted otherwise, all analyses were based on responses from the last four trials of each day, and all estimates of asymptotic performance were derived from the final two blocks of 40 trials each. To determine whether an asymptote had been reached, the mean difference between these trial blocks was tested for each group, but the test is mentioned explicitly only when significant.[4] Since $\Pi_{11} = \Pi_{22}$ and $\Pi_{21} = \Pi_{12}$ in this study, more reliable estimates of conditional response probability were obtained by calculating $p_{11}(n,t)$, an average of the estimates $p_{11}(n,t)$ and $p_{22}(n,t)$, and by calculating $p_{1j}(n,t)$, an average of the estimates $p_{21}(n,t)$ and $p_{12}(n,t)$.

[4] In this paper the .05 level is adopted as the minimum level for significance.

TABLE 1

TERMINAL ESTIMATES OF CONDITIONAL RESPONSE PROBABILITY

Data	Intertrial Interval							
	8 Sec.		30 Sec.		3 Min.		60 Min.	
	Rep.[a]	Alt.[b]	Rep.	Alt.	Rep.	Alt.	Rep.	Alt.
All *Ss*	.73[c]	.85	.71	.52	.59	.50	.51	.50
Nonfixated *Ss*	.74[c]	.89	.74	.66	.70	.57	.57	—
Nonfixated *Ss*, 2nd trial only	.68	.94	.69	.67	.66	.60	.49	—

[a] Probability measure is $p_{11}(\infty, t)$.
[b] Probability measure is $p_{12}(\infty, t)$.
[c] The difference between the final two trial blocks was significant ($P < .001$).

During the first 40 trials, conditional response probability clustered about .50 for all experimental groups, except Groups 8-A and 30-A. As will be seen, the corresponding estimates of .74 and .63 for Groups 8-A and 30-A were probably inflated by alternation tendencies. All experimental groups, except 8-R, appeared to reach an asymptote between 80 and 160 trials. No relationship was detected between rate of asymptotic approach and the experimental conditions. The above trends also characterize the performance of the subclass of "nonfixated" *Ss*, to be defined later.

Terminal estimates of conditional response probability for all experimental *Ss* are given in Table 1. Even though the estimate for Group 8-R jumped from .69 to .77 during the final two trial blocks, it is retained as the best available estimate of the asymptote for that group.

Since, for all groups beyond the 8-sec. interval, the distributions associated with the terminal estimates were highly skewed with heterogeneous variances, the main effects were tested by applying the extended median test for k independent samples (Siegel, 1956). The significant intertrial interval effect ($P < .001$) was interpreted as qualitative evidence in favor of the theory. The significant reward schedule effect ($P < .01$) indicated that conditional response probability was higher under the repetition schedule than under the alternation schedule. However, the significance of these effects, particularly that of the latter, must be viewed cautiously since the interaction probably was great.

According to statistical learning theory (Estes, 1954), when $\Pi = .50$, as it does for all *Ss* in this study, individual estimates of $p_1(\infty)$, the asymptotic probability of an E_1 response, should approximate a binomial distribution about .50. Instead, the obtained dispersion described a tri-modal curve with slightly more than 50% of the estimates deviating beyond the 99% acceptance interval for .50 (see Table 2). Approximately

TABLE 2

TERMINAL ESTIMATES OF $p_1(\infty)$ VALUES FOR ALL *Ss*

$p_1(\infty)$	No. of *Ss*	$p_1(\infty)$	No. of *Ss*
.00	18	.56–.66	20
.01–.11	0	.67–.77	14
.12–.22	3	.78–.88	4
.23–.33	5	.89–.99	6
.34–.44	23	1.00	29
.45–.55	37		

TABLE 3

FREQUENCY OF FIXATED *Ss*
(Total number of *Ss* in parentheses)

Reward Schedule	Intertrial Interval			
	8 Sec.	30 Sec.	3 Min.	60 Min.
Repetition	1 (17)	2 (17)	8 (15)	12 (15)
Alternation	2 (18)	17 (20)	19 (20)	18 (18)

Note.—Of the 19 *Ss* in the control group, 3 were fixated.

58% of the 159 *Ss* had $p_1(\infty)$ values larger than .50. According to the binomial test, this split was significant ($P < .05$), indicating an over-all group preference for the left side of the maze.

Several trends among the *Ss* that deviated beyond the 99% acceptance interval for .50, henceforth called "fixated" *Ss*, warrant mention. Most position habits appeared early in the study: 50% of the fixated *Ss* developed position habits during the first 40 trials; 75% during the first 80 trials; and 95% during the first 160 trials. According to the chi-square test, position preferences occurred most frequently among *Ss* under the alternation schedule ($P < .001$) and under highly distributed trials ($P < .001$) (see Table 3).

The fixated *Ss* complicate any quantitative test of the trace-fluctuation model since conditional response probability is suppressed toward .50 whenever strong position preferences occur. Consequently, if the model were to be tested fairly, its application was necessarily limited to *Ss* having no pronounced position habits. To insure this, all fixated *Ss* were excluded from the theoretical analysis.[5] Although this exclusion eliminated more than one-half of the *Ss*, it yielded adjusted mean estimates of $p_1(\infty)$ which differed only slightly from .50. Furthermore, it led to sets of individual estimates which approximated normal distributions, permitting the use of parametric statistics.

For the nonfixated *Ss* in the control group, estimates of $p_{ij}(n,t)$ were employed as measures of reward alternation tendencies, and estimates of $p^{A}{}_{ij}(n,t)$, the probability of an E_j response after Trial n, given an E_i response on Trial n, were employed as measures of response alternation tendencies. When successive rewards are independent, the theory predicts that $p_{ij}(n,t)$ should approach Π minus a small correction term and that $p^{A}{}_{ij}(n,t)$ should approach Π minus an indeterminate, but presumably small, correction term. The terminal estimate for $p_{ij}(n,t)$ was .72, and that for $p^{A}{}_{ij}(n,t)$ was .63. Both estimates differed significantly ($P < .001$) from Π, suggesting the presence of strong reward and response alternation tendencies.

The frequency with which the control *Ss* alternated relative to both the previous reward and response remained fairly constant throughout the experiment. To determine the relationship between terminal estimates of reward and response alternation tendencies, the correlation ratio, *eta*, was calculated. The obtained value of .14 differed insignificantly from zero; no consistent relationship existed between the terminal estimates of reward and response alternation for individual *Ss*.

The terminal estimates of conditional response probability for the nonfixated experimental *Ss* are given in Table 1. The variances associated with these estimates were .009, .008, .019, .003, .004, and .005 for Groups 8-R, 30-R, 180-R, 3600-R, 8-A, and 30-A, respectively. Bartlett's test gave a nonsignificant result when applied to this data. The variance model used in the two-

[5] To facilitate comparison, the three fixated *Ss* in the control group were also excluded from subsequent analyses.

way analysis is the one which can be applied when data are missing completely from a given subclass (Snedecor, 1946).

The main effects need not be qualified since the interaction was not significant. Once again, the significant intertrial interval effect ($P < .001$) was in the direction predicted by the theory. The largest decrease in conditional response probability occurred between the 8- and 30-sec. intervals ($P < .02$). None of the other decreases between adjacent intervals was significant.

The significant reward schedule effect ($P < .01$) was now in the direction anticipated by the experimental design, i.e., conditional response probability was higher under the alternation schedule than under the repetition schedule. At the 8-sec. interval, conditional response probability under the alternation schedule was significantly higher ($P < .001$) than that under the repetition schedule. None of the differences at either the 30-sec. or the 3-min. intervals was significant.

The significant difference at the 8-sec. interval probably can be attributed to the inflationary and depressive effects of alternation tendencies on $p_{ij}(\infty,t)$ and $p_{ii}(\infty,t)$, respectively. It was assumed that any alternation tendencies would have the same magnitude of effect on both estimates. Since these effects would be in opposite directions, $p_{ix}(n,t)$, an *unbiased* estimate of the probability of an E_i response following an E_i occurrence, could be evaluated by averaging estimates of $p_{ii}(n,t)$ and $p_{ij}(n,t)$ obtained under the repetition and alternation schedules.

A more rigorous test of the model required that Equations 5 and 6 be averaged to obtain an expression for $p_{ix}(\infty,t)$. If, when the sequential dependencies between rewards are symmetrical, it is assumed that $j_1 = j_2 = j_i$ and $N_1 = N_2 = N_i$, the resulting expression for $p_{ix}(\infty,t)$ has the same form as Equations 5 and 6, except for the disappearance of the correction terms.

Given the above conditions, one could estimate the theoretical parameters j_i and N_i from the data at two intervals and generate exact predictions for $p_{ix}(\infty,t)$ at the remaining two intervals. After $j_i = .01$ and $N_e/N_i = .22$ had been estimated from the 8-sec. and 3-min. intervals, predictions were generated for the 30-sec. and 1-hr. intervals. As can be seen in Table 4, the discrepancies between obtained and predicted values were rather large.

To test these discrepancies, weighted means were substituted for the unweighted estimates of $p_{ix}(\infty,t)$. Since the revised estimates differed only slightly from those appearing in Table 4, the previously estimated values of j_i and N_e/N_i were retained. Although still large, the differences between obtained and predicted values did not lead to significant t ratios.

The slow decay rate of the trace stimuli, as suggested by the estimate of .01 for j_i, implies that traces from several previous rewards may have been present on the 3rd, 4th, and 5th trials for each day. To partial

TABLE 4

OBTAINED AND PREDICTED CONDITIONAL RESPONSE PROBABILITIES FOR THE NONFIXATED *Ss*

Data	$p_{ix}(\infty,t)$	Intertrial Interval			
		8 Sec.	30 Sec.	3 Min.	60 Min.
Responses on Trials 2–5	Obt.	.82[a]	.70	.64[a]	.57[b]
	Pred.	.82	.80	.64	.50
Responses on Trial 2	Obt.	.81[a]	.68	.63[a]	.49[b]
	Pred.	.81	.79	.63	.50

[a] Data used to estimate theoretical parameters.
[b] Weighted mean.

out these effects, the responses of the nonfixated *Ss* on the second trial were analyzed for the last 20 days of the experiment. (This period corresponded to the final 80 trials.) None of the previously nonfixated *Ss* fell in the fixated category as a result of this reanalysis. The revised terminal estimates for the nonfixated *Ss* are listed in Table 1.

After $j_i = .01$ and $N_e/N_i = .27$ had been estimated from the revised data at the 8-sec. and 3-min. intervals, predictions of conditional response probability were made again for the 30-sec. and 1-hr. intervals. Although the discrepancy between obtained and predicted values at the 30-sec. interval remains large, that at the 1-hr. interval is now smaller, as can be seen in Table 4.

The model can be extended to account for the presence of more than one trace. For subsequences of similar events under the repetition schedule, the necessary modification of Equation 5 reads

$$p_{ii}(\infty,t,x) = \frac{1}{N_e + \sum_{k=0}^{x-1} N_i(t+kc)} \{[\Pi + \sum_{k=0}^{x-1} \theta(1-\Pi)(1-\theta)^k]N_e + [\Pi_{ii} + \sum_{k=0}^{x-1} \theta\Pi_{ii}(1-\Pi_{ii})(1-\theta\Pi_{ii})^k] \sum_{k=0}^{x-1} N_i(t+kc)\} \quad [7]$$

where c equals the time required for a complete trial cycle and x equals the number of available traces on any given trial. According to Equation 7, as the number of available traces increases, conditional response probability approaches Π_{ii} plus a growing, but still small, correction term.

For subsequences of alternating events under the alternation schedule, the necessary revision of Equation 6 reads

$$p_{ij}(\infty,t,x) = \frac{1}{N_e + \sum_{k=0}^{y-1} N_i(t+2kc) + \sum_{m=0}^{z-1} N_j[t+(1+2m)c]} \times \{[\Pi - \sum_{k=0}^{y-1} \theta\Pi(1-\theta)^k + \sum_{m=0}^{z-1} \theta(1-\Pi)(1-\theta)^m]N_e + [\Pi_{ij} - \sum_{k=0}^{y-1} \theta\Pi_{ij}(1-\Pi_{ij})(1-\theta+\theta\Pi_{ij})^k] \sum_{k=0}^{y-1} N_i(t+2kc) + [\Pi_{ii} + \sum_{m=0}^{z-1} \theta\Pi_{ii}(1-\Pi_{ii})(1-\theta\Pi_{ii})^m] \sum_{m=0}^{z-1} N_j[t+(1+2m)c]\} \quad [8]$$

where, when $x = 1$, Equation 6 is appropriate; when x is odd and greater than one, $y = \frac{1}{2}(x+1)$ and $z = \frac{1}{2}(x-1)$; and when x is even, $y = z = \frac{1}{2}x$. According to Equations 6 and 8, conditional response probability is a cyclical function of the number of available traces.[6]

To test the predictions of Equations 7 and 8, the small correction terms were ignored, and c was set equal to the intertrial interval plus 8, where the latter quantity is the average

[6] Straightforward modifications of Equations 7 and 8 yield predictions for conditional response probability following any subsequence of events. In this study, however, no attempt was made to check the theory against the remaining subsequences since they occurred infrequently.

TABLE 5

OBTAINED AND PREDICTED CONDITIONAL RESPONSE PROBABILITIES FOR THE NONFIXATED *Ss* AS A FUNCTION OF THE NUMBER OF PREVIOUS REWARDS ON A GIVEN DAY DURING THE LAST 20 DAYS OF THE EXPERIMENT

Group	Probability Measure	Previous Rewards			
	$p_{11}(\infty, t)$	E_1	E_1E_1	$E_1E_1E_1$	$E_1E_1E_1E_1$
8-R	Obt.	.68[a]	.81	.82	.84
	Pred.	.68	.75	.78	.80
30-R	Obt.	.69	.82	.82	.83
	Pred.	.68	.74	.78	.80
	$p_{21}(\infty, t)$	E_1	E_1E_1	$E_1E_1E_1$	$E_1E_1E_1E_1$
8-A	Obt.	.94	.91	.91	.86
	Pred.	.90	.50	.68	.54

[a] Data used to estimate theoretical parameters.

time in seconds required by *Ss* to run from the choice point to the goal box, and to eat the mash reward. Predictions were limited to the comparatively large number of nonfixated *Ss* in Groups 8-R, 30-R, and 8-A.

The estimates $j_i = .001$ and $N_e/N_i = 1.21$ for Equation 7 were obtained from the revised terminal estimates for Groups 8-R and 180-R, shown at the bottom of Table 1. Predicted and obtained values for Groups 8-R and 30-R are given in Table 5 as a function of the number of previous rewards on a given day during the last 20 days of the experiment. (It should be noted that the number of observations on which the estimates in Table 5 were based decreased as the subsequence became longer. Thus, for Group 8-R, the estimate following E_i was based on 320 observations, while the remaining estimates following E_iE_i, $E_iE_iE_i$, and $E_iE_iE_iE_i$ were based on 286, 256, and 227 observations. The number of observations decreased at approximately the same rate for Groups 30-R and 8-A. Except for slight variations, all *Ss* within a group contributed an equal number of observations to each of the estimates for that group.) Although the predicted values consistently underestimated the obtained values for Groups 8-R and 30-R in Table 5, the predicted trend occurred, and most of the discrepancies were fairly small, suggesting the presence of traces from at least two, and perhaps as many as four, preceding rewards.[7]

Estimates of j_i and N_e/N_i could not be made for Equation 8 since the revised terminal estimate of .94 for Group 8-A was larger than the predicted asymptote. Nevertheless, to provide a rough indication of the predictions generated by Equation 8, it was assumed that $j_i = .001$ and $N_e/N_i = .01$. As can be seen in Table 5, not only are the discrepancies between obtained and "predicted" values very large, but the predicted trend is not apparent.

Finally, the theory can be checked against an additional data point at the 18- to 22-hr. interval. When the terminal estimates of conditional response probability were derived from responses on the first trials of the last 20 days of the experiment, only 25 of the 140 experimental *Ss* fell in the nonfixated category. The terminal estimate of .53 for these *Ss* did not differ significantly from the predicted value of approximately .50.

DISCUSSION

In spite of precautions taken to eliminate preferential cues or treatment in the T maze, the *Ss* collectively demonstrated a significant preference for the left side of the maze. Nevertheless, the high incidence of fixated *Ss* probably cannot be

[7] Since all of these discrepancies were in the same direction as the growing correction term, the fit between theory and data for Groups 8-R and 30-R is slightly closer than indicated in Table 5.

attributed solely to peculiarities of the experimental situation, since 30 *Ss* had position habits toward the right side of the maze. These results, when coupled with similar findings reported by Estes (cited by Bitterman, Wodinsky, & Candland, 1958) for rats and by Anderson and Grant (1957) for humans, suggest that the prediction of statistical learning theory for individual variability may need revamping.

Why did the incidence of fixated *Ss* rise with increases in trial spacing? Perhaps strong response and reward alternation tendencies interfered with the development of position habits. On the basis of previous studies (Dember & Fowler, 1958), it is reasonable to assume that the frequency of alternation decreased as trial spacing became more distributed. Consequently, the development of position habits was no longer obstructed at the larger intervals.

Why did the fixated *Ss* under the alternation schedule outnumber those under the repetition schedule? The *Ss* with position habits under the alternation schedule received continuous partial reinforcement, while their counterparts under the repetition schedule received alternate periods of regular reinforcement and nonreinforcement. Presumably then, the position habits of the *Ss* under the repetition schedule should have been more susceptible to extinction. According to this explanation, the responses of these *Ss* should lag behind shifts in reinforcement from one side of the maze to the other, while the previously reinforced position habit was being extinguished. This lag tended to occur among the nonfixated *Ss* under the repetition schedule.

In accordance with the theory, conditional response probability approached .50 as trial spacing increased. On the other hand, exact predictions of conditional response probability, particularly at the 30-sec. interval, did not accurately reflect the obtained values. Conceivably, much of the discrepancy between theory and data could have been caused by the selective elimination of fixated *Ss* as the intertrial interval increased. To avoid selectivity, an alternative design could be employed. Instead of being assigned permanently to a given interval, *Ss* could be rotated through all intervals after reaching the "asymptotic" series; even if it were necessary to make a separate analysis based on the nonfixated *Ss*, the same *Ss* would now be represented at all intervals.

The agreement between obtained and predicted conditional response probabilities, as a function of the number of previous rewards, was fairly good for Groups 8-R and 30-R, but poor for Group 8-A. Although no attempt was made to determine precisely whether only two, or as many as four, preceding rewards affected the responses of the *Ss* under the repetition schedule, the present results resemble Anderson's (1956) findings that the responses of human *Ss* were affected by three preceding rewards, and for at least one group, by four preceding rewards. For most groups in the study by Anderson, the mean probability of a repetition response, as estimated from the last 100 trials of a 300-trial acquisition series, was an increasing function of the length of subsequences of similar events. As noted by Anderson, however, the form of the function could have been caused by a growing correction term rather than by an accumulation of traces, since the slope of the function was slight in most cases. In the same study, terminal behavior following subsequences of alternating events reflected neither the cyclical trend specified by Equation 8 nor the performance of *Ss* in the present study. More specifically, for all groups except one, the mean probability of an alternation response increased as a function of subsequence length. Although at odds, the findings of both studies suggest that Equation 8 may be inadequate, but before discarding it, at least the finding for Group 8-A should be confirmed at a few larger intervals where the confounding effect of strong alternation tendencies should be negligible. Such a confirmation was not attempted in this study due to the high incidence of fixated *Ss* in Groups 30-A, 180-A, and 3600-A.

SUMMARY

According to the trace-fluctuation model, terminal estimates of conditional response probability should match the sequential dependencies between reinforcements only when trial spacing is massed. When trial spacing is distributed, these estimates should approach a value of approximately .50.

The *Ss* were 159 rats, each of which received from 200 to 400 trials in a simple T maze. Each *S* was assigned permanently to one of four intertrial intervals: 8 sec., 30 sec., 3 min., or 1 hr. To adjust for possible alternation tendencies, one-half of the experimental *Ss* worked under a "repetition" reward schedule, while the other half worked under an "alternation" reward schedule. Also a control group worked under a sequentially independent reward schedule.

As trial spacing increased, terminal estimates of conditional response probability for all experimental *Ss* approached .50.

More than 50% of all *Ss* developed position preferences. These preferences occurred most frequently among *Ss* working under the alternation schedule at the larger intertrial intervals. All *Ss* with strong position habits were excluded from subsequent analyses.

Alternation occurred with respect to both the previous response and the previous reward. No consistent relationship existed between these tendencies for individual *Ss* in the control group.

As trial spacing increased, terminal estimates of conditional response probability for all *Ss* without position habits also approached .50. Exact predictions of conditional response probability did not accurately reflect the obtained values, particularly at the 30-sec. interval.

A reanalysis of the data suggested that *Ss* under the repetition schedule were responding in terms of traces from at least two, and perhaps as many as four, preceding rewards. The agreement between theory and reanalysed data was fairly good for selected groups under the repetition schedule, but poor for *Ss* under the alternation schedule at the 8-sec. interval.

REFERENCES

ANDERSON, N. H. Effect of first-order conditional probability in a two-choice learning situation. Unpublished doctor's dissertation, Univer. of Wisconsin, 1956.

ANDERSON, N. H., & GRANT, D. A. A test of a statistical learning theory model for two-choice behavior with double stimulus events. *J. exp. Psychol.*, 1957, **54**, 305–317.

BITTERMAN, M. E., WODINSKY, J., & CANDLAND, D. K. Some comparative psychology. *Amer. J. Psychol.*, 1958, **71**, 94–110.

BOTTENBERG, R. A. Asymptotic level, shift rapidity and conditional relationships in a two-choice learning situation. Unpublished doctor's dissertation, Stanford Univer., 1956.

BURKE, C. J., & ESTES, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

DEMBER, W. N., & FOWLER, H. Spontaneous alternation behavior. *Psychol. Bull.*, 1958, **55**, 412–428.

ENGLER, J. Marginal and conditional stimulus and response probabilities in verbal conditioning. *J. exp. Psychol.*, 1958, **55**, 303–317.

ESTES, W. K. Individual behavior in uncertain situations: An interpretation in terms of statistical association theory. In R. A. Thrall, C. H. Coombs, & R. L. Davis (Eds.), *Decision processes*. New York: Wiley, 1954. Pp. 141–149.

ESTES, W. K. Statistical theory of spontaneous recovery and regression. *Psychol. Rev.*, 1955, **62**, 145–154.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GOODNOW, J. J., & POSTMAN, L. Probability learning in a problem solving situation. *J. exp. Psychol.*, 1955, **49**, 16–22.

HAKE, H. W., & HYMAN, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

SNEDECOR, G. W. *Statistical methods*. (4th ed.) Ames: Iowa State Coll., 1946.

WITTE, R. S. A stimulus-trace hypothesis for statistical learning theory. Unpublished doctor's dissertation, Stanford Univer., 1958.

(Received June 23, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

THE RELATION BETWEEN MEAN REWARD AND MEAN REINFORCEMENT

ALLAN M. LEVENTHAL,[1] RICHARD F. MORRELL,[2] ELMER F. MORGAN, JR.,[3] AND CHARLES C. PERKINS, JR.

Kent State University

A number of experiments have been conducted to determine the effects that variations in magnitude of reward or in percentages of responses rewarded have on rate of learning and final level of performance. Several experimenters have studied the interaction of these variables with human *S*s (Edwards, 1954; 1956). However, there are no reports of such experiments with animal *S*s although the value of such studies has been pointed out by Jenkins and Stanley (1950, p. 200).

The general problem with which we are concerned is the question of whether or not rats' responses can be predicted from the assumption that following extensive training behavior will be of whatever sort maximizes reward over a series of trials.

Precise definitions of critical terms will be presented now to facilitate description of the specific problems investigated and discussion of the results obtained.

The term *reward* refers to a goal object and magnitude of reward is some physical measure of the amount of the goal substance (e.g., grams of food). *Mean reward* is the average amount of the goal object presented per response including nonrewarded responses. If only certain response occurrences are rewarded and the reward is the same following all rewarded responses, *magnitude of reward* refers to its magnitude on those trials which are rewarded. Thus, with 50% of the responses rewarded (ordinarily described as 50% reinforcement) mean reward is half the magnitude of reward.

Reinforcement is the strengthening effect of stimuli on antecedent responses. A more precise statement becomes somewhat involved. Consider two mutually exclusive choice responses, R_A and R_B, which would occur equally often (within chance limits) had they been regularly followed by the same event. If both responses occur many times to the same stimulus, each being followed by a different event, and following this training R_A occurs regularly while R_B no longer occurs, then the event (X_A) which follows R_A is (by definition) more reinforcing than X_B. Similarly, if the events following one or both of the responses differ from one occurrence of a response to the next and R_A comes to occur regularly (or predominantly), the *mean reinforcement* following R_A is greater than the mean reinforcement following R_B.

The present study was designed to determine which of two responses, if either, will become dominant after extensive training when mean reward is the same following both responses but magnitude of reward, and thus percentage of responses rewarded, is different. Each group of *S*s received one unit of reward in one arm of a maze on every run to this side; on the other side they received two units

[1] Now with V. A. Mental Hygiene Clinic, Des Moines, Iowa.

[2] Now rehabilitation counselor, Ohio State Bureau of Vocational Rehabilitation.

[3] Now County School Diagnostician, Monroe County, Michigan.

TABLE 1

SUMMARY OF DIFFERENCES IN PROCEDURES FOR THE FOUR GROUPS

Group	*E*	Maze	*Ss*		Daily Trial Block			Reward Unit
			N	Sex	2	4	8	
FS	AML	A	12	F	—	Da. 1–10	Da. 11–25	.25 gm.
FL	RFM	B	14	M	Da. 1–2	Da. 3–36	—	1.0 gm.
WS	EFM	C	12	F	—	Da. 1–28	—	2 drops
WL	EFM	C	12	F	—	Da. 1–28	—	10 drops

of reward on half the runs and no reward on the others. Two groups ran under water deprivation for water reward and two under food deprivation for food reward. Mean reward differed for the two groups run under each deprivation condition. If mean reward is the sole determiner of mean reinforcement, no significant preference for either response would be acquired by *Ss* in any group.

METHOD

Since the work reported here was done as part of three separate M.A. theses, all groups were not run by the same *E*, and certain characteristics of apparatus and *Ss* were not exactly the same. These differences, which are summarized in Table 1, are assumed to have had no measurable effect on the general nature of the results. Thus, for the most part, this report is in the form that would be used if all *Ss* had been run under conditions that were identical except for deprivation and reward conditions.

Subjects.—Data are reported from 50 naive albino rats from the colony maintained by the Kent State University Psychology Department. Age of *Ss* was 83–180 days at the beginning of training. The distribution of the *Ss* in the different groups is summarized in Table 1. Seven additional *Ss* were discarded from Group FL owing to errors on the part of *E*. Their performances were typical up to the time they were discarded. One *S* was discarded from Group WS when it failed to meet a predetermined criterion of making a choice within 20 min. subsequent to Day 4.

Apparatus.—Three single choice point E-shaped alley mazes were used. All mazes had distinctively different goal boxes, and the stem and each arm were painted differently. Alleys in each maze were 4.5 in. wide and 3.75 in. deep; sides and floor were wood, and the roof was hardware cloth. The mazes differed with respect to length of sections, style of painting, and floor covering. Details are given in the theses by Leventhal (1955), Morgan (1957), and Morrell (1957).

Lighting was provided by a single bulb above the median plane of each maze. A one-way vision screen was placed between *E* and the maze.

Procedure.—There were features common to the treatment of all groups, namely: (*a*) every *S* received a reward on 100% of his runs to one arm of the maze and twice as large a reward on 50% of his runs to the other arm; (*b*) rewarded and nonrewarded runs to the 50% side were irregular in sequence within the limitation that each day (or when *Ss* received only two trials a day, within successive blocks of four trials) half the runs to the 50% side were rewarded; (*c*) after *S* had run to one side on half its trials for that day, the door on that side of the choice point was closed forcing *S* to the opposite side (thus, the number of runs to each side was equal at the end of each day's training); (*d*) the 100%, 1 unit of reward condition was on the right for just half the *Ss* in each group, to counterbalance for possible position or color preference; (*e*) within each group, time in the goal box was the same on every trial. Table 1 summarizes the differences between the groups with respect to the major variable, reward condition.

At least eight days prior to the beginning of training the groups run to food reward were placed on a schedule of limited feedings every 24 hr. and reduced to 80% ad libitum body weight. This regimen was maintained throughout the experiment. Food rewards were pellets of *S*'s standard diet—Purina Laboratory Chow. The water deprivation groups had access to water only on rewarded trials and during the first hour after the completion of each day's training. The

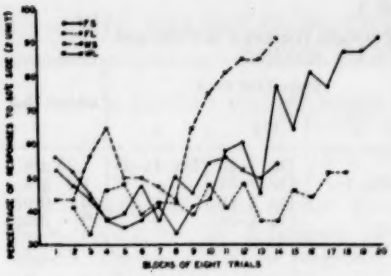

FIG. 1. Learning curves for the four groups. Percentages of the first half of the day's runs which were to the 50%, 2-unit side are plotted for each block of eight trials (including all runs) for each group.

intertrial interval was not controlled precisely but ran between 5 and 25 min. Its average duration probably was longest for Group FL and shortest for Group FS.

Six of the *Ss* in Group FL were held in a delay box for 30 sec. between the choice response and entrance to the goal box on every trial. Since the choices of these *Ss* did not differ systematically from those of the other *Ss* in this group, the delay was assumed to have no effect on the measure obtained and results of delay and nondelay *Ss* in Group FL were pooled.

At all times water was available to *Ss* in Groups FS and FL, except that *Ss* in Group FS did not have access to water when they were in the maze. Dry food was available to *Ss* in Groups WL and WS except when they were in the maze.

As indicated in Table 1, the total number of trials was 160 and 144 for Groups FS and FL, respectively, while the water groups were only run for 112 trials. It was originally planned to run them longer, but training was discontinued because a bitter substance had inadvertently been placed in the water given to some *Ss* on several trials starting with Trial 113.

RESULTS AND DISCUSSION

Throughout, the data presented are for the first half of the runs on each day. Data based on all free trials lead to the same conclusions and thus are not presented.

Learning curves for each group are presented in Fig. 1 where percentage of the first half of the day's runs to the 50%, 2-unit side are plotted as a function of blocks of 8 trials (including both free and forced runs). Examination of Fig. 1 shows that the two small-reward groups acquired a marked tendency to run to the 50%, 2-unit side and that the two large-reward groups developed no clear preference for either side.

The differences between the learning curves for the two small reward groups is of little importance as (*a*) several factors which might affect learning rate differed systematically between the two groups, and (*b*) the magnitude of food reward comparable to a given amount of water is not known. However, it is of interest that for both these groups the preference for the 50%, 2-unit side begins to show up only after considerable training.

In Groups FS and WS the number of *Ss* running to the 50%, 2-unit side on a majority of their last 12 trials (including only the first half of the trials on each day) was 11 and 12, respectively. Expansion of the binomial indicates that these preferences are significant at beyond the 1% level. In each of Groups FS and WS, 11 of the 12 *Ss* ran to the 50%, 2-unit side more times in their last 12 trials than on their first 12, indicating that the increase in runs to the 50%, 2-unit side cannot be attributed entirely to chance. Comparable tests for the other two groups do not approach significance. Presumably, even if a larger number of *Ss* or more extensive training should disclose a significant preference for the 50%, 2-unit side with these particular magnitudes, a preference for the 100%, 1-unit side would be obtained if absolute magnitude were large enough so that consumption of one unit would produce satiation.

As measured by Fisher's exact test (Fisher, 1946, p. 96) the proportion of *Ss* making a majority of their last 12 runs to the 50%, 2-unit side was reliably different for the two food groups ($P < .03$) and for the two water groups ($P < .005$). Thus, for both food and water reward there is an interaction between mean reward and the relative reinforcing properties of two units of reward for 50% of the responses vs. one unit for 100%.

No other trends (e.g., preference for the 100%, 1-unit side on the part of Groups FS, FL, and WL early in training) are significant.

The explanation of the preference for 50%, 2 units over 100%, 1 unit when the unit is small cannot be specified with assurance. However, the explanation which seems most reasonable is that complex internal responses which presumably accompany consumption of food and water cannot be entirely appropriate if the goal object is quite small. That is, only if the magnitude of reward given a rat is greater than .25 gm. of food would the concomitant responses associated with consumption and digestion of food closely approximate the optimum level. The observation that when they have free access to food and water, rats ordinarily consume moderately large quantitites whenever they consume any at all is in line with this interpretation.

It appears probable that further related experimental work which is integrated by a concept such as mean reinforcement can serve as the basis for construction of equal unit scales of the magnitude of the reinforcing effect of various food or water rewards. Since magnitude of reward and reinforcement magnitude apparently are not equivalent, construction of equal unit scales of magnitude of reinforcement would be helpful, perhaps even necessary, for the development of a precise quantitative behavior theory.

Summary

Four groups of rats were run in single choice mazes where they received one unit of reward for each run to one side and two units for 50% of the runs to the other. Number of occurrences of each response was equated by forced trials. The groups differed in reward condition with large and small reward groups for both food and water.

The two small reward groups, for which .25 gm. of food or 2 drops of water was equal to one unit, acquired a significant preference for the response followed by the 50%, 2-unit reward condition. The two large reward groups did not acquire a preference for either response.

The results indicate that response selection is not determined entirely by the mean reward following choice responses.

References

Edwards, W. The theory of decision making. *Psychol. Bull.*, 1954, **51**, 380–417.

Edwards, W. Reward probability, amount, and information as determiners of sequential two-alternative decisions. *J. exp. Psychol.*, 1956, **52**, 177–188.

Fisher, R. A. *Statistical methods for research workers.* (10th Ed.) London: Oliver and Boyd, 1946.

Jenkins, W. O., & Stanley, J. C., Jr. Partial reinforcement: A review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.

Leventhal, A. M. The relative strength of responses followed by 100% reinforcement of one unit of food versus 50% reinforcement of two units. Unpublished master's thesis, Kent State Univer., 1955.

Morgan, E. F., Jr. Strength of responses followed by different percentages and magnitudes of water reward. Unpublished master's thesis, Kent State Univer., 1957.

Morrell, R. F. The relative reinforcement provided by one gram of food 100% of the time versus two grams of food 50% of the time. Unpublished master's thesis, Kent State Univer., 1957.

(Received May 16, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

A STUDY OF SEMANTIC GENERALIZATION THROUGH THE USE OF ESTABLISHED CONCEPTUAL MEDIATIONS [1]

RICHARD J. KAPLAN [2]

University of California, Los Angeles

The study of semantic generalization can be carried out in either one of two ways. Jeffrey and Kaplan (1957) have attacked the problem by building up the associations upon which the generalization is based in the laboratory through the use of nonsense syllables. Bugelski and Scharlock (1952) used the same method in the study of mediated association. Others (Foley and Cofer, 1943; Razran, 1949) have studied semantic generalization in which the requisite associations were assumed to be already present in *S* by virtue of his previous experience.

The same hypothesis is being tested in either case: that a generalizing or mediating response is elicited due to some previous associations on the part of *S* and that these produce facilitation in learning during the critical portion of the experimental situation. The two methods differ only in their technique.

It is the purpose of this investigation to combine the two techniques in a single experiment so that the relative power of each may be evaluated.

[1] The present study is based on a dissertation submitted to the Graduate Division of the University of California at Los Angeles in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author is indebted to Wendell E. Jeffrey, Chairman of the Doctoral Committee, and to John P. Seward and Irving Maltzman for their guidance during the course of the research.

[2] Now at Litton Industries, Beverly Hills, California.

METHOD

Learning material.—In an attempt to develop a set of norms of common English words to be used in the study of concept formation, Underwood and Richardson (1956) obtained the association responses of college sophomores to about 200 words. The normative group was trained to give only responses in terms of a sense quality. The learning materials for the present experiment was taken entirely from Underwood and Richardson's list of norms. These lists were constructed giving the percentage of a particular response given to a word by the group as a whole, and if two words elicited the same response in a high percentage of the cases, they are considered by the writer to have preformed associations on the basis of the common response word.

For example, the word ROUND was given as a response to the word WHEEL in 94% of the cases and was given as a response to the word GLOBE in 95% of the cases. The pair WHEEL–GLOBE then, is called a "mediation pair" in the context of this experiment, by virtue of the fact that the same response word, i.e., ROUND, was elicited by both members of the pair in a great proportion of the cases. No indication can be gotten from the lists, however, as to whether WHEEL and GLOBE are directly associated, and if so, to what extent this association exists. Pairs of this type are considered to have preformed associations, as opposed to post-formed associations which are induced in the experimental situation as described later. The words pre- and post-formed are referenced to the point in time at which the experiment begins.

The pairs of words which appear in the test list are divided into two classes defined as mediation and non-mediation pairs on the basis of the value of the association as measured by the Underwood and Richardson lists. These classes were set up using two criteria. First, the mediation pairs are composed of words which had elicited a common response in more than 80% of the *Ss*. Second, the non-mediation pairs are

TABLE 1

PARADIGMS FOR LEARNING

Treatment	List 1	List 2	List 3
E_1	B-A (WHEEL—ROUND)	C-A (GLOBE—ROUND)	B-C (WHEEL—GLOBE)
E_2	B-a (IVY—BATON)	C-a (SPINACH—BATON)	B-C (IVY—SPINACH)
E_3	b-a (COAL—BUCKLE)	c-a (HALLWAY—BUCKLE)	b-c (COAL—HALLWAY)
C_1	B-d (MILK—CORN)	C-e (CHALK—CAVE)	B-C (MILK—CHALK)
C_2	b-d (ANCHOR—COFFEE)	c-e (GINGER—ICICLE)	b-c (ANCHOR—GINGER)

composed of words which had not elicited a common response in more than 10% of the *S*s. The further restriction was added that none of the words in the non-mediation pairs could have elicited in more than 10% of the *S*s the response which is used to define any one of the mediation pairs.

Experimental design.—The experiment consists of five treatments placed in what Lindquist (1953) calls a mixed design. This design permitted the use of each *S* as his own control in that he receives all of the treatments in the experiment. Thus, the treatment effects are within-*S* effects. The between-*S* effects will be discussed as the design is developed.

The treatments are concerned with different aspects of the semantic generalization problem and consist of three experimental and two control groups as follows: E_1—Mediation pairs with the actual preformed association used in the training series; E_2—Mediation pairs with a common response in the training series but not the actual preformed association; E_3—Non-mediation pairs with a common response used in the training series; C_1—Mediation pairs with no common response in the training series; C_2—Non-mediation pairs with no common response in the training series, i.e., no possibility for semantic generalization to occur.

The paradigm for the learning of the three lists is given in Table 1 for each treatment. The following notation is used in the table: *B* and *C* each stand for one member of a mediation pair; *b* and *c* each stand for one word of a non-mediation pair; *A* is the actual response word used in determining a mediation pair; *a* is a common response between a pair; and *d* and *e* are different responses to a pair. An example of each of the pairs is presented under each of the symbolic representations.

Each list is composed of three words for each of the treatment groups so that each *S* learned a 15-pair list composed of all the treatments. In order to counteract the confounding of the treatment effects with the difficulty of the words it would be necessary to rotate the words so that all of them appeared in each group. Because of the insufficient length of the Underwood and Richardson list, this was not entirely possible, but an approximation to it was made. The mediation pair treatments, E_1, E_2, and C_1, were arranged in three groups so that each pair appeared in each of the treatments. For each of these three alternations the two treatment effects involving non-mediation pairs were also switched around, making a total of six groups.

Six *S*s were used in each of the groups in the experiment, making a total of 36 for the whole experiment. These were further divided into two groups, men and women, adding one more variable to the experimental design.

Apparatus and procedure.—The paired-associate words used as the learning material for the experiment were typed on adding-machine tapes. The stimulus word alone was presented first, and then the stimulus word appeared again along with the response word with which it was to be associated. Three different random orders were used. A Gerbands M-1 memory drum, modified to accept a list of this size, was used. The drum revolved one space every 3 sec., and one blank space occurred between the end of

one presentation of the complete list and the beginning of the next.

The *Ss* were drawn from the students enrolled in an introductory psychology course. Each *S* appeared for the experiment in one 2-hr. session during which he learned each of the three lists to the criterion of three perfect repetitions plus two additional repetitions whether perfect or not.

The *S* was seated before a cardboard screen which separated him from *E* and the major part of the apparatus. The face of the memory drum projected through a 3 × 5 in. hole in the center of the screen about 3 in. below the eye level of the seated *S*.

After instructions for the experiment were read to him, *S* was allowed to ask any questions regarding the procedure, and *E* made certain that *S* understood what he was to do. After the first list was started, no questions regarding the experiment were answered, and *E* did not communicate with *S* other than to give the brief instructions regarding the rest periods and subsequent lists.

After the completion of the final list *S* was interrogated in an informal and unstructured manner allowing him to provide all the information he could without any direct questions from *E*. The purpose of this was to determine the extent to which the *Ss* realized the purpose of the experiment and could document this with specific recollection of the actual learning material. Because of the possibility of communication to future *Ss*, the full purpose of the experiment was not revealed to any *S*.

For each *S* a record was made of the performance at each trial, a trial being the possibility of making an anticipation. The *S*'s performance was recorded as correct, no response, or the word said if an incorrect word were given in anticipation.

Results and Discussion

The data were collected in terms of the number of errors on the third or test list, an error being defined as either no response or an incorrect anticipation to the stimulus word when it appeared in the window of the drum.

The means and *SD*'s of errors for all *S*'s in each of the treatment groups are presented in Table 2.

The design of the experiment calls for the data to be analyzed in terms of an analysis of variance. This Treatments × Subjects design assumes that the interaction effects are distributed with homogeneous variance among the treatments. A Bartlett test for homogeneity of variance (Edwards, 1950) was run by computing an interaction effect for each score in the table and testing the five treatment variances of these interactions for homogeneity. The resulting χ^2 was 23.6 ($P < .001$, 4 *df*), indicating heterogeneity of variance. Inspection of Table 2 shows that one of the variances, namely that of the E_1 treatment, differs markedly from the others. Both logarithmic and square-root transformations of the data failed to yield homogeneity of variance. On the basis of this result it was decided to drop the E_1 treatment from the analysis. When this was done, the Bartlett test χ^2 was 6.6, which for 3 *df* is nonsignificant. The analysis of variance was performed after exclusion of the E_1 treatment, and is presented in Table 3.

TABLE 2

Means and *SD*s of Errors by All *Ss* in Treatment Effects on the Test List

Measure	Treatment				
	E_1	E_2	E_3	C_1	C_2
Mean	3.58	6.17	10.89	6.89	10.97
SD	1.20	4.77	5.68	3.39	5.42

The proper error term for tests of the "within" effects is the $MS_{T\times S}$. When these tests are made, the only significant effect is seen to be the Treatments main effect, and this, highly so. Since none of the other interactions is significant they may be pooled with the T × S interaction to give a better error term with additional 33 *df*. This was not considered necessary, however, owing to the

TABLE 3

ANALYSIS OF VARIANCE OF TEST LIST DATA

Source	*df*	*MS*	*F*
Within *Ss*	(108)	(201.45)	
Treatments	3	230.72	15.57*
T × G	15	14.60	—
T × S	72	14.82	—
T × Sex	3	21.17	1.43
T × G × Sex	15	8.95	—
Between *Ss*	(35)	(96.60)	—
Groups	5	78.03	—
Ss	24	99.08	—
Sex	1	128.44	1.30
G × Sex	5	96.84	—
Total	143		

* $P < .001$.

already highly significant F ratio. The proper error term to be used in testing the "between" effects is the S main effect, and it is readily seen that none of these is significant.

The critical difference between means for $P = .05$ was calculated to be 1.54. Reference to Table 2 will indicate those treatments which differ significantly.

From the differences between the treatment means a hierarchical order from greater to lesser generalization can be formed among the four treatments. This order is $E_2 - C_1 - E_3 - C_2$, with the two differences at either end of the series, $E_2 - C_1$ and $E_3 - C_2$ being the only nonsignificant ones. The questioning of *Ss* at the end of the experiment yielded the information that the recall of the pairs followed the same hierarchical order as the generalization as derived from the differences between the treatment means.

Since the E_1 treatment was not included in the analysis it must be interpreted separately from the rest of the data. Inspection of Table 2 shows that E_1 was much easier than any of the others. This is further emphasized by the fact that the F ratio between E_1 and E_2, the computation of which does not assume homogeneity of variance, turns out to be 15.80. This is highly significant. Twenty-eight out of the 36 *Ss* made no errors after the first presentation of the test list, whereas only 10 got perfect scores on the E_2 part. This result is to be expected since the basis for generalization between the pairs in this treatment comes not only from the preformed association but also from the reinforcement of the postformed association during the learning by the presence of the mediating response in the training list.

What effect the presence of the preformed associations might have had on the learning of the other words in the list cannot be determined from the data. It may be that the presence of the common association between the stimulus and the response in the E_1 part of the list led S to look for similar connections and thus produced an added "instruction" to him not intended by E. A suggestion offered to counteract this in further research of this type is to eliminate the E_1 group from the list under the assumption that this type of learning is different from, or at least of a different order of magnitude from, the others and cannot be studied in the same situation.

The necessity for performing an analysis of covariance on the data might have arisen due to the differential effect of reinforcement in the different treatments. The Pearson r between the number of errors on the two training lists and the treatment was .59. When, however, the same coefficient was computed with the E_1 treatment omitted, r was only .29. This indicated that most of the differential reinforcement occurred in the E_1 treatment, and that with the omission of the E_1 treatment, the analysis of covariance is no longer required.

It may be seen from Table 2 that the lowest two groups with respect to learning were the ones involving the non-

mediation pairs. The words in these two groups never appeared in the part of the list involving mediation pairs, and it might be said that the differences here are due to differences in the ease of learning the particular words rather than to the treatment effects themselves. This confounding was unavoidable because of the inadequate length of the norms for constructing the lists for this experiment. Two factors, however, argue against this interpretation. The first is the insignificant group main effect and T × G interaction in the analysis, which shows that in whatever arrangement the words occur, there is no difference. The second comes from an estimation of the relative familiarity of the *S*s with the words taken from the norms constructed by Thorndike and Lorge (1944). The use of the "T" scale shows that the mediation pairs had an average occurrence of 218 per million and the non-mediation pairs, 207 per million. This difference does not seem sufficient to account for the highly significant differences between the treatment means.

The most obvious result in the experiment is that generalization occurred only in the two parts of the test list involving mediation pairs, since only these two differed significantly from the control part. This is interpreted as indicating that preformed associations are a more powerful factor in producing generalization than postformed associations.

Although the difference between the E_2 and C_1 treatments is considerably larger than that between the E_3 and C_2 treatments, it still is not significant. This emphasizes the fact that the postformed associations produce no generalization effect in this particular situation. Not only do the postformed associations fail to show any generalization when they are acting alone, but also, they fail to increase the effect of the preformed associations when they are put in combination with these. Even though this difference between the E_2 and C_1 treatments is not significant, the fact that it is not negative is of some importance. Jeffrey and Kaplan (1957) found that generalization and not interference could explain the results of their study, and the same interpretation is invoked to explain these results. Although the addition of the postformed associations does not augment the generalization effect in the mediation pairs, neither does it diminish this effect by interfering with the learning process.

The comparison of the $C_1 - C_2$ difference and the $E_3 - C_2$ difference shows that in the total situation the preformed association alone produces generalization whereas the postformed association alone does not.

Summary

Thirty-six college students learned three lists of paired words to mastery. The lists were divided into five treatment effects representing two control groups and three experimental groups. The experimental groups offered the possibility for semantic generalization to occur by virtue of one of two processes or by a combination of both of these. An estimation of the preformed associations was taken from norms developed on college sophomores by Underwood and Richardson. The postformed associations were built up by associating each word of the final pair to be tested to a common response word. The measure of generalization in the experiment was the facilitation in the learning of the test list produced by either of the two processes. The measure of learning was the number of errors made by the subject in anticipating the correct response word.

The results of the experiment showed the following: (1) Postformed associations alone did not produce generalization between words in the test list. (2) Preformed associations alone did produce generalization between words in the test list. (3) The combination of postformed association with preformed association did not produce an increment in generalization over and above the effect of preformed association alone. (4) The recall of the subjects with respect to pairs of words on the test list and the training which led up to them, followed exactly the same order as the ease of generalization.

It was concluded from these results that preformed associations are more effective in producing generalization than postformed associations. It was concluded also that the effect of the postformed associations was

not an inhibitory one, in that the presence of these with preformed associations did not make the combination harder to learn than preformed associations alone.

References

BUGELSKI, B. R., & SCHARLOCK, D. P. An experimental demonstration of unconscious mediated association. *J. exp. Psychol.*, 1952, **44**, 334–338.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

FOLEY, J. P., & COFER, C. N. Mediated generalization and the interpretation of verbal behavior: II. Experimental study of certain homophone and synonym gradients. *J. exp. Psychol.*, 1943, **32**, 169–175.

JEFFREY, W. E., & KAPLAN, R. J. Semantic generalization with experimentally induced associations. *J. exp. Psychol.*, 1957, **54**, 336–338.

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

RAZRAN, G. Semantic and phonetographic generalization of salivary conditioning to verbal stimuli. *J. exp. Psychol.*, 1949, **46**, 337–367.

THORNDIKE, E. L., & LORGE, I. *The teacher's book of 30,000 words.* New York: Columbia Univer. Press, 1944.

UNDERWOOD, B. J., & RICHARDSON, J. Some verbal materials for the study of concept formation. *Psychol. Bull.*, 1956, **53**, 84–95.

(Received May 23, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

COMPARISON OF PREDICTIONS AND ESTIMATES IN A PROBABILITY LEARNING SITUATION[1]

E. D. NEIMARK

New York University

AND

E. H. SHUFORD

Psychometric Laboratory, University of North Carolina

Shuford (1957) has shown that when *S*s are presented with a 20 × 20 matrix of elements, a proportion, Π, of which are in one state (e.g., blue squares, vertical bars) and the remaining proportion, 1 − Π, are in another state (e.g., red squares, horizontal bars), *S*s are remarkably accurate in estimating the proportion of either class of element. With 400 elements presented for a 1-sec. interval, there is no possibility that *S*s are counting. Since *S*s estimate so accurately when the entire population of elements is present, the question arises as to whether *S*s will be equally accurate in their estimates if they sample a population of elements sequentially and without replacement. Such a situation is essentially that used by Humphreys (1939) and others (Estes & Straughan, 1954; Grant, Hake, & Hornseth, 1951; Neimark, 1956) for probability learning where *S*s are asked to predict which one of two possible events will occur on the next trial.

It is commonly found in these probability learning tasks that the final proportion of predictions of a given class of events tends to approximate the relative frequency of occurrence of that class. This being the case, one would expect that direct estimates of relative frequency as a response measure should resemble proportion of predictions. However, on a given trial *S* either predicts the event or he does not; in order to transform predictions into estimates (or response probabilities) it is necessary either to average across a block of trials for each *S*, or to average across *S*s on a given trial. In either case, the grain of the resulting scale is a function of the number of trials or *S*s used and one encounters all the hazards of averaging data (Bakan, 1954; Estes, 1956; Sidman, 1952).

The purposes of this study were (*a*) to obtain direct estimates of event relative frequency, (*b*) to compare these estimates with relative frequency of prediction of the event, and (*c*) to determine how both response measures are affected by availability of information concerning the preceding trials.

[1] This study was done in support of ARDC Project 7740 while both authors were at the Operator Laboratory, Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole and in part by or for the United States Government. This study has not received the review and approval normally given to this Center's reports. The views, conclusions, and recommendations expressed or implied herein do not necessarily reflect the official views or policies of the United States Air Force or of the Air Force Personnel and Training Research Center. The authors wish to thank Earl F. Lingle for his assistance in running the *S*s and analyzing the data.

METHOD

Subjects.—The *S*s were 72 Air Force recruits. All *S*s had an Armed Forces Qualification Test score of 65 or above, which is roughly equivalent to a Stanford-Binet IQ of 108 or more (Gragg & Douglass, 1956). Their modal age was 17.

Apparatus.—Three decks of one hundred 3 × 5-in. index cards were stacked face down on a modified desk calendar stand. Each card had one of two letters printed upon it. A cover card atop each deck contained the two letters which would appear in that deck; the letter whose frequency was to be estimated was underlined. The letters of the alphabet were paired randomly without replacement. Three sequences of card orders were drawn randomly with $\Pi = .67$; the actual relative frequencies of the underlined letter in the three decks were .67, .63, and .65. The within-deck sequences and order of decks were the same for all *S*s.

Procedure.[2]—The *S* was seated at a desk containing the three-deck stack of cards, three mimeographed answer sheets, and a pencil. He was told by *E* that he had three well-shuffled decks of cards, each card of which contained one of the two letters on the cover card for that deck, and that his task was to determine what percentage of the cards in the deck contained the letter underlined on the top card. At the beginning of each deck, *S* was required to write down an initial estimate of the percentage of cards in the deck which would contain the underlined letter. On each trial for 100 trials per deck, *S* wrote in the first column of the answer sheet his prediction of the letter which would appear on the card to be turned up and, after turning the card, he copied in the second column the letter which appeared. Two groups of 18 *S*s each, the PI and PC groups, filled in only these two columns of the answer sheet for each card. Two other groups of 18 *S*s each, the PEI and PEC groups, were required in addition to write in a third column their estimate of the percentage of cards in the deck which would contain the underlined letter. In other words, *S*s in these groups made a new estimate on each trial in addition to the initial estimate which was required of all *S*s. After completing the first deck, all four groups of *S*s went directly on to the second and third decks, each of which used a different pair of letters.

[2] The study presented is part of a larger experimental design which, because of a variety of misfortunes, was never completed. One-half of the *S*s, the PC and PEC groups, each received a large index card with which to cover the preceding predictions and outcomes (and estimates in the case of the PEC group). The remaining groups, PI and PEI, used no cover cards and could refer back to preceding trials. Each of the four groups was run in two sessions of 9 *S*s each; each session required about an hour.

RESULTS AND DISCUSSION

The mean relative frequency of prediction of E_1 (the more frequent, i.e., underlined, letter) on successive 10-trial blocks is given in Fig. 1. Data are plotted separately for each of the four groups and three decks. The data differ from the results of most studies employing this type of procedure in two respects: (*a*) with the exception of the last deck the curves do not start at a chance level; (*b*) predictions for the PE groups do not level off at Π as those for the P groups do. Analysis of variance of E_1 predictions on the last 30 trials of each deck shows that the frequency of E_1 prediction is significantly higher for the PE groups than for the P groups ($F = 11.10$, 1 and 68 *df*). Although there is a good deal of within-group variability, the mean relative frequency of E_1 for the P groups approximates Π whereas the entire distribution is shifted upward in the case of the PE groups with a mean exceeding Π. Neither information, nor deck effects, nor any interaction is significant at the .05 level. Evidently, the availability of information concerning preceding events has no effect on the final level of mean E_1 prediction.

When mean *estimates* of E_1 (Fig. 2) are compared with proportion of E_1 *predictions* (Fig. 1), it is apparent that estimates also tend to start out at a higher than chance level. The final levels of E_1 estimates appear to approximate Π. Comparisons of final

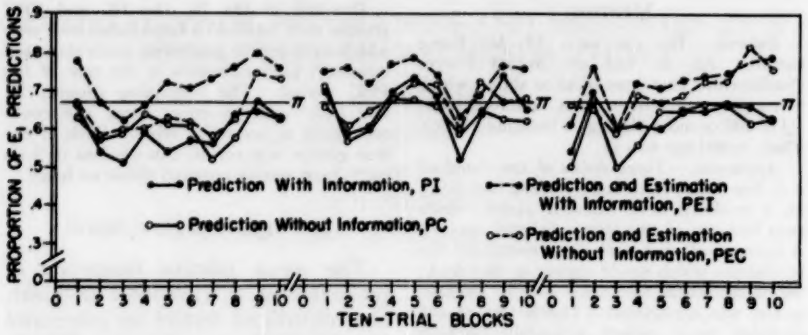
PROPORTION OF E1 PREDICTIONS
Prediction With Information, PI
Prediction Without Information, PC
Prediction and Estimation With Information, PEI
Prediction and Estimation Without Information, PEC
TEN-TRIAL BLOCKS

FIG. 1. Proportion of E_1 predictions in successive 10-card blocks for each of the four groups on each of three decks.

level of mean E_1 estimates with final level of E_1 predictions on the last 30 trials of each deck indicate that: (*a*) predictions are consistently higher than estimates for the PE groups ($F = 6.87$, 1 and 68 *df*); whereas, (*b*) predictions for the P groups are not significantly different from estimates for the PE groups ($F = 2.08$, 1 and 68 *df*). Although comparison among final levels for the PE groups suggests that predictions are higher and estimates are lower (i.e., more accurate) when information concerning the past events is available, neither information effects nor interaction effects is significant at the .05 level.

Comparison of successive E_1 predictions for the P groups with successive E_1 estimates for the PE groups (Fig. 2) supports the suggestion from comparisons of final levels that predictions and estimates are comparable response measures in this situation. Data on groups giving estimates only (EI and EC) would, of course, provide a stronger basis for this conclusion and it is unfortunate that these data are not available. However, data on EI groups from similar experiments show that *Ss* follow the sequence quite closely and suggest that data for an EI group would differ from that for the PEI group only in decreased within-group variability. Although estimates and predictions appear to be comparable measures it is clear that *Ss*

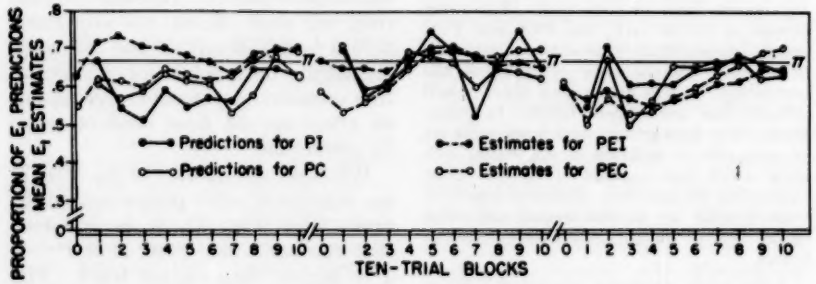

FIG. 2. Comparison of proportion of E_1 estimates for the P groups with mean E_1 estimates of the PE groups on successive 10-card blocks of the three decks.

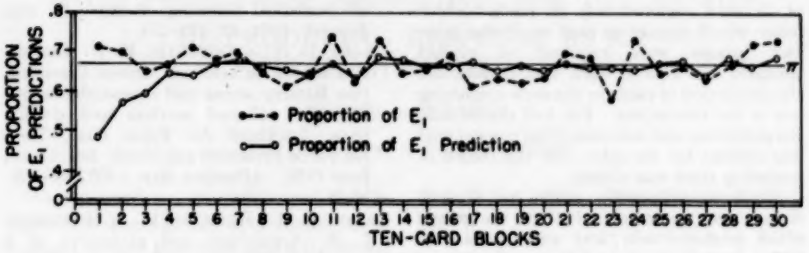
PROPORTION OF E_1 PREDICTIONS
.8
.7
.6
.5
.4
0
Proportion of E_1
Proportion of E_1 Prediction
Π
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
TEN-CARD BLOCKS

FIG. 3. Proportion of E_1 predictions on successive 10-card blocks in relation to actual E_1 proportion. Data are based upon an independent group of 198 *Ss*; actual E_1 proportion represents the mean of nine random sequences with $\Pi = .67$.

in the usual probability learning situation are not all estimating the relative frequency of E_1 before making predictions, for when *Ss* are required to estimate as well as predict E_1, the frequency of prediction is significantly higher than when estimates are not required.

Initial E_1 estimates and predictions.— It will be remembered that both estimates and predictions of E_1 started at an above-chance level. In order to determine whether probability learning curves starting from a chance level could be obtained in this situation, 198 *Ss* were run under a modified PC procedure in which initial estimates were not required. Nine random sequences of 300 cards were drawn with $\Pi = .67$. Twenty-two *Ss* were run on each deck. For eleven of these *Ss* the letter which occurred with probability .67, E_1, was underlined on the top card; for the other eleven *Ss* the letter which occurred with probability .33, E_2, was underlined. Proportion of E_1 prediction was tabulated for all *Ss*. The results are plotted in Fig. 3 along with the actual proportion of E_1's in the nine random sequences (broken line). The curve starts at a chance level and shows a negatively accelerated increase to an asymptote about equal to .67. Clearly the traditional probability learning curves can be obtained in this experimental situation.

Our procedure in the main experiment differed from the traditional procedure in that all *Ss* were required to give an initial estimate of E_1 frequency at the start of each deck. We had assumed that the distribution of initial estimates (Bayes distribution[3]) would be symmetrical about a mean of .50, but the obtained distributions were negatively skewed with means of .57, .63, and .60 for the three decks. Thus, it would appear that underlining one of two letters and requiring *S* to estimate its relative frequency at the start of the experiment tends to increase his initial estimate. Moreover, it has already been shown that requiring *S* to estimate E_1 throughout the deck raises the level of concomitant E_1 prediction. Hence, the initial estimate alone should raise the initial level of E_1 prediction. When no additional estimates are required, the effect of the initial estimate should decay on successive trials so that the final level of E_1 prediction is not affected. This effect is borne out in the data for the PI and PC groups.

SUMMARY

A factorial design was used in a probability learning situation to investigate the effects of (*a*) response required and (*b*) information available. Four groups of 18 *Ss* each went through three decks of 100 cards, each card of which contained one of two letters. The three 100-card sequences were randomly drawn with a probability of .67 for one of the letters and .33 for the other. Two groups

[3] Investigation of the effect of manipulation of initial Bayes distribution has been undertaken by Shuford and Land, who have developed a model for the prediction of successive estimates in this situation (Shuford, 1957).

of *Ss* were required only to predict which letter would appear on each card; the other two groups were required to predict similarly and also on each trial to estimate the proportion of cards in the deck containing one of the two letters. For half the *Ss* only the prediction and outcome of the current trial was visible; for the other half the record of preceding trials was visible.

Prediction proportions on the last 30 cards of each deck approximated .67 for the groups which predicted only, and was significantly higher for the groups which predicted and estimated. Estimates for the latter two groups were similar to the proportions of predictions of the prediction-only groups. Deviations from theoretical values for both initial and final levels of prediction were discussed in terms of the effects of estimates upon predictions. Availability of information had no significant effect upon predictions or estimates.

References

BAKAN, D. A generalization of Sidman's results on group and individual functions and a criterion. *Psychol. Bull.*, 1954, **51**, 63–64.

ESTES, W. K. The problem of inference from curves based on group data. *Psychol. Bull.*, 1956, **53**, 134–140.

ESTES, W. K., & STRAUGHAN, J. H. Analysis of a verbal conditioning situation in terms of statistical learning theory. *J. exp. Psychol.*, 1954, **47**, 225–234.

GRAGG, D. A., & DOUGLASS, H. J. Conversion tables for selected Airman Classification Battery scores and comparable scores on other selected services and civilian tests. Lackland Air Force Base, Tex.: Air Force Personnel and Train. Res. Cent., June 1956. (*Develpm. Rep. AFPTRC-TN-56-89.*)

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42**, 1–5.

HUMPHREYS, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.

NEIMARK, E. D. Effects of type of nonreinforcement and number of alternative responses in two verbal conditioning situations. *J. exp. Psychol.*, 1956, **52**, 209–220.

SHUFORD, E. H. The estimation of proportions in two binary situations. Paper read at Amer. Psychol. Ass., New York, September, 1957.

SIDMAN, M. A note on functional relations obtained from group data. *Psychol. Bull.*, 1952, **49**, 263–269.

(Received May 26, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

SOME CONDITIONS GOVERNING THE USE OF THE CUE-PRODUCING RESPONSE AS AN EXPLANATORY DEVICE[1]

JOSEPH DE RIVERA

U. S. Naval School of Aviation Medicine

Two similar stimuli are made more distinguishable if a response is learned to one stimulus and a different response is learned to the other stimulus. It has been suggested (James, 1950; Miller & Dollard, 1941) that the different responses add different stimuli to the original stimuli thus making the original stimuli more disparate. This concept of cue-producing responses leading to an acquired distinctiveness of the stimuli to which they are attached has been fruitfully used as an explanatory concept by Dollard and Miller (1950). While in general the above concept has been useful, in work with rats by Lawrence (1949), with children by Cantor (1955), and with adults by Gagne and Baker (1950), some recent evidence suggests that the manner in which the concept is applied must be defined more closely.

Robinson (1955) had one group of *Ss* learn a different name to each of 10 different fingerprints, and another group of *Ss* learn one name to five of the prints and a second name to the other five prints. Assuming that each of these responses furnished *S* with additional cues, we might expect the first group to have acquired a distinctiveness of the fingerprints while the second group acquired a similarity of the fingerprints. However, when *Ss* were given a task involving print discrimination, both groups performed equally well.

Goss (1953) had some *Ss* learn names to four different light brightnesses while a control group was simply told to pay attention to the lights. After this training, *Ss* learned to select a different switch to each light brightness. Only those *Ss* with 100% overlearning of the names performed better than the control group on the second task.

In a recent review of the literature on stimulus predifferentiation, Arnoult (1957) concluded from the available data that: "*Positive transfer from stimulus predifferentiation training may be expected after a minimum of 4 to 8 pretraining trials and reaches a maximum after 8 to 12 pretraining trials*" (p. 341). Arnoult further concluded that a satisfactory explanation for the positive transfer obtained from predifferentiation training would be developed only when the physical and experimental aspects of the stimulus could be better described.

A preliminary experiment by the present author suggested one way in which the conception of the stimulus should be refined in order to apply the concept of the cue-producing response to successfully explain the above-mentioned data. The *Ss* first learned letters, then numbers, to a set of fingerprints. Paralleling Robinson's results, those *Ss* who learned two letters to 10 prints showed as much transfer as those who learned 10 letters. However, many *Ss* who had learned 10 letters stated they had tried to forget the letters and memorize the numbers directly to the prints. Furthermore, when *Ss* who had learned two letters indicated the characteristics of the prints to which they had memorized these two letters, it became apparent that most of them had actually learned 10 responses. That is, they had singled out 10 different characteristics, one from each print, and had memorized a letter to each of these characteristics. Instead of learning that all five prints called by the same letter showed, perhaps, pronounced loops, *S* would learn that one print had a whorl in the lower left corner, another had a slash in the middle, and so on.

The above experiments led to two hypotheses as to the manner in which the concept of cue-producing responses should be applied. Let *overt*

[1] Opinions or conclusions contained in this report are those of the author. They are not to be construed as necessarily reflecting the view or the endorsement of the Navy Department.

responses be defined as those responses which *E* requires *S* to give when different stimuli are presented, and let *perceptual responses* be defined as those responses which constitute *S*'s representation of the stimulus. In the following experiment the perceptual responses are inferred to be the selective seeing of certain characteristics of the fingerprint stimuli, and are represented verbally by such statements as "a whorl in the center," "a very dark print," or "wavey lines." It is then hypothesized that: (*a*) overt responses learned to the stimuli will not be used to furnish cues for additional learning unless these overt responses are strongly attached (overlearned) to the stimuli; and (*b*) in cases where the overt responses are not overlearned and hence not used to furnish cues, perceptual responses learned to the stimuli will be used as cues for further learning, and hence, the number and kind of these perceptual responses will be more important than the number and kind of overt responses.

To test the first hypothesis *S*s were given either 8 or 16 trials to learn to write 10 letters in response to 10 fingerprints. Then *S*s were given 8 trials to learn 10 numbers to the same print. The *S*s who learned letters for 16 trials should more frequently use these letters as cues for learning numbers to the prints. Having these additional cues available, they should show more positive transfer.

To test the second hypothesis, *S*s were given 8 trials to learn two letters to the 10 prints. Some *S*s were told to learn each letter to the characteristics of each individual print. Their perceptual responses should be similar to those of the *S*s who learned 10 letters to the individual prints. Another group of *S*s was told to memorize each of the two letters to the characteristics which were possessed in common by the prints with the same letter. Their perceptual responses should be different from those of the other *S*s. The second hypothesis predicts that, although only two overt responses were learned, those *S*s who learned 10 perceptual responses should perform as well on the transfer task as *S*s who learned (but did not overlearn) 10 overt responses. On the other hand, those *S*s who learned only two perceptual responses should not have as many cues available and hence should not transfer as well.

METHOD

Subjects.—Groups of naval aviation cadets were tested during their first week of preflight training at Pensacola, Florida. The size of the groups varied from 19 to 36 men. The *S*s were told that the test was for research purposes only.

Procedure.—A set of 10 different fingerprints reproduced on slides served as stimuli for all the groups tested. On the first trial a slide was projected for 1 sec., then its correct response (a letter or number) was announced and the slide remained on the screen for two more seconds. Immediately afterwards, the next slide of the set was shown. After this introductory trial, *S*s were given 8 (or 16) training trials. On these trials a slide was shown for 4 sec., during which time *S*s were required to guess and write down the slide's correct number. After this time the correct number was announced; in the next 2 sec. *S* marked his guess right or wrong, and had a chance to associate the correct number with the slide, then the slide was removed. (In order to detect any tendency for *S* to write the numbers immediately after they were announced, on Trials 4 and 8 the numbers were not announced until all 10 slides had been shown.) Four seconds after a slide was removed, a new slide was presented until all 10 fingerprints had been shown. The *S*s then counted their correct guesses and wrote down the sum. After 30 sec. the same 10 slides, mixed up in a different order, were again presented until *S*s had completed 8 (or 16) trials. The order in which the slides were presented was initially determined randomly.

Initial training.—Group A was given 8 trials to learn 10 letters (A–J) to the set of 10 fingerprints. On Trial 8 an average of 7.8 letters reported correctly indicated a substantial amount of learning had occurred.

Group B was given 16 trials to learn 10 letters (A–J) to the set of 10 prints.

Group C was given 8 trials to learn two letters to the set of 10 prints. Half of the prints were arbitrarily called "C" and half were called "G." The *Ss* in Group C were told ". . . there is nothing in common between the five slides called 'G.' I just arbitrarily picked five of them and called them 'G.' Thus, when you memorize the slides don't waste your time looking for common characteristics—you'll just get mixed up. Instead, start right away to memorize to each individual slide its correct letter."

Group D was given 8 trials to learn two letters to the set of 10 prints. The same prints which had been arbitrarily called "C" and "G" for Group C were again called "C" and "G." However, *Ss* in Group D were told, ". . . five of these slides have certain aspects in common and are called 'C,' the other five slides also have certain characteristics in common and are called 'G.' It is your job to learn to call each slide by its correct letter. Obviously, the best way to do this is to spot the common characteristics of the 'C' slides and the 'G' slides."

Group E received no initial training.

Transfer training.—Approximately 5 min. after the initial training, all five groups learned a new set of responses, the numbers 1–10, to the same set of fingerprints used in the preliminary training. Each *S* received eight trials on this new task. After these trials *Ss* were asked to give the characteristics of the slides which they had used in memorizing the responses.

Results

Initial training.—The mean numbers of errors made by each group in the first eight training trials are as follows: Group A—27.6 (SD = 13.0); Group B—30.2 (SD = 14.3); Group C—19.5 (SD = 8.3); Group D—18.7 (SD = 8.7).

As was expected, the means and SD's of Groups A and B (which had to give 10 responses) are much larger than the means and SD's of Groups C and D (which had to give only two responses). As is proper, however, the mean and SD of Group A are not significantly different from those of Group B; and the mean and SD of Group C are not significantly different from those of Group D.

Stimulus characteristics important in transfer training.—After the eight transfer learning trials, *Ss* in Groups C and D were asked to write or draw the characteristics of the different fingerprints to which they had memorized the responses "C" and "G." Analysis of these characteristics showed that 29 of the 36 *Ss* in Group C gave a different characteristic for each slide. As instructed, most of Group C had memorized the two responses to 10 different slide characteristics. In Group D, 26 of the 30 *Ss* gave characteristics which were common to the "C" slides or to the "G" slides. For example, one *S* selected the fact that "almost all C's were lighter," and another remarked, "most C's had narrow loops." As instructed, most *Ss* in Group D had memorized the two responses to characteristics common to the slides with the same letter.

The *Ss* in Groups A and B were asked to indicate for each number (1–10) whether they had memorized the number to the previously learned letter or directly to some characteristic of the print. If to the letter, they were asked to write the letter down beside the number. A frequency distribution shows that *Ss* in Group A used a letter to mediate 39% of their transfer trials. The *Ss* in Group B used a letter to mediate 60% of their transfer trials.

Since the distributions did not approximate normality, a Mann-Whitney U test was run. The differences are significant at the .05 level and part of Hypothesis 1 is thus confirmed. Many *Ss* in Group B made

TABLE 1

MEANS, *SD*'S AND *t* TESTS OF ERRORS MADE IN TRANSFER TRAINING

Group	*N*	Mean	*SD*	Diff.	*t*	*P* (One-Tail)
A	19	16.63	6.63	A–B	1.81	.04
B	34	12.35	8.79			
C	36	16.39	12.67	C–D	1.78	.04
D	30	22.13	13.13			
E	84	28.26	12.00	D–E	2.32	.02

statements such as, "At first I got the number from the letter; gradually I grew to recognize the number from the print." Several *Ss* indicated the mnemonic device they used to learn the number to the letter such as, "the word 60 for six-d."

Transfer training.—The mean number of correct responses made on each trial by the various groups during transfer training is shown in Fig. 1 and the total number of errors made by each group is shown in Table 1. As predicted by Hypothesis 1, Group B made fewer errors on the transfer task than did Group A. The difference is significant at the .04 level (one-tail).

In order to establish that the superiority of *Ss* in Group B was due to their greater utilization of the letters as cues, *Ss* in both Groups A and B were divided at the median into those who reported using many and those who reported using few letters, and into those who made many and those who made few errors, and a chi-square test was run. Those *Ss* who more frequently used the letters as cues tended to make less errors on the transfer task ($P = .04$, one-tailed). However, inspection of the data showed that *Ss* who used the letters tended to have made fewer errors when they initially learned the letters to the prints.

An analysis of covariance was used in order to eliminate the effects of this ability difference. The *Ss* were placed in a 2×2 table according to whether they had used more or less than the median number of letters, and according to whether they had received 8 or 16 trials in which to learn the letters. Quenouille's (1948) technique was used to correct for unequal *N*'s, and the sums of squares were adjusted to allow for differences in initial ability. The interaction term was tested by using the within-cells mean squares as an error term. An $F = .62$ was obtained. Since the interaction was insignificant, the interaction and within-cells mean squares were pooled to test the row and column effects. The column (number of trials) effect was significant at the .03 level (one-tailed). The row (use of letter as cue) effect was insignificant ($F = .68$). Thus, when account is taken of the greater learning ability of those *Ss* who use letters as cues, the use of letters does not significantly improve *S*'s transfer learning.

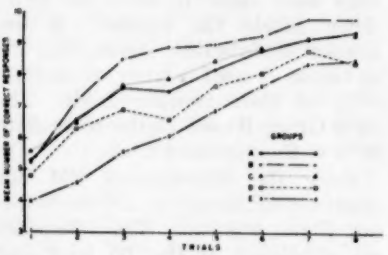

FIG. 1. Mean number of correct responses on each trial of transfer training.

As predicted by Hypothesis 2, Groups A and C made approximately the same number of errors, while both groups made less errors than did Group D. The difference between Groups A and C is negligible. The difference between Groups C and D is significant at the .04 level (one-tail). The difference between Groups A and D fails to reach the .05 level of significance when the unequal variances are equalized by the use of a Behrens-Fisher test. The evidence

cannot be called conclusive but clearly supports Hypothesis 2.

The transfer task variances of Groups C, D, and E are approximately equal to each other and to the initial task variances of Groups A and B. However, all five of these variances are far greater than the transfer task variances of Groups A and B. The differences are all significant between the .05 and .01 level (two-tails). Evidently, practice in learning 10 different overt responses to the prints significantly reduced the variance of the distribution of errors made in the later learning of 10 responses to the prints. Practice in learning two different overt responses to the prints did not lead to this effect.

Discussion

We tentatively conclude that the extent to which overt responses, learned to stimuli in preliminary training, will be used to mediate the learning of new responses, is a function of how well the overt responses are learned to the initial stimuli. Thus, Group B used the responses to a greater extent than Group A. However, *S*s who reported using these mediating responses seemed to have no advantage. The superiority of Group B must be attributed to some other factor. Perhaps *S*s in Group A incurred more negative transfer because they received an equal amount of training on both the initial and transfer tasks. Perhaps, even when an overt response was not reported as the primary cue, it still served to help differentiate the print by becoming part of a complex stimulus pattern, and thus the degree to which the response was learned was important.

When an overt response is not well learned to a stimulus it will not function as a primary cue for later learning. However, the necessity for an overt response will cause *S* to learn several other responses incidentally. These other responses include adjustive responses, such as when and how to look for the stimulus and when and how to give the required response, and perceptual responses such as "one slide has a loop in the center," another has "straight lines on the bottom," and so forth. The perceptual responses may themselves furnish cues for future learning, even though the initial overt responses which occasioned the perceptual responses to be made have been forgotten. Therefore, in studying the acquired distinctiveness of cues we must be sophisticated as to just which responses are furnishing additional cues. In the experiment reported above, when the overt responses (letters) were not well attached to the stimuli they did not usually function as cues. However, the groups learning different perceptual responses (Groups A and C) showed more transfer than the group learning similar perceptual responses (Group D). Therefore, the perceptual responses did function as cues and their number and similarity were more important than the number and similarity of the overt responses. This interpretation also may account for the results of Robinson (1955). We need simply postulate that his group which learned two overt responses to fingerprints actually learned 10 perceptual responses as well, as did groups in our preliminary studies.

It is interesting to note that when we view the transfer resulting from stimulus predifferentiation in this second way, that is, as a function of the perceptual responses learned rather than the overt responses taught, we are in effect saying that the labels taught *S* are irrelevant and merely force him to search for discriminable aspects of the stimuli. This, however, is the "attention to cues" hypothesis mentioned by Arnoult (1957) and favored by Hake and Eriksen (1955) as an explanation for the positive transfer resulting from stimulus predifferentiation. We may, therefore, deduce this point of view from a consideration of the general cue-producing response.

While we can say that perceptual cues probably played a role in the discriminability of our fingerprints, we cannot say

whether the distinctive perceptual responses aided discrimination or the common perceptual responses hindered discrimination or, as is most likely, both of these effects occurred. To state the direction of the effect another group would be necessary in order to control for the practice effects which were probably one reason the groups with preliminary training learned the numbers with greater ease than the group which had no preliminary training. In other words, we cannot say here how much of our effect is due to the adaptive responses we mentioned above. It should also be noted that *S*s in Group D learned the general observing response of looking for common rather than discrete characteristics in the fingerprints; this habit might account for part of their failure to transfer as much as the other groups.

Summary

Conflicting studies on the effect which teaching differential responses to cues has on the acquired distinctiveness of these cues led to the following two hypotheses: (*a*) Distinctive overt responses attached to stimuli will not be used to furnish additional cues for later learning unless the responses are strongly attached (overlearned) to the stimuli. (*b*) When overt responses are not strongly attached to the stimuli, the perceptual responses made in the attempt to learn the overt responses furnish cues for later learning to the stimuli and hence are more important than the overt responses in determining future transfer.

To test these hypotheses one group of *S*s (Group A) learned distinctive overt responses to a set of stimuli; Group B overlearned the same distinctive overt responses to the stimuli; Group C learned equivalent overt responses but distinctive perceptual responses to the stimuli; Group D learned equivalent overt responses and equivalent perceptual responses to the stimuli; and Group E learned no responses to the stimuli. Immediately thereafter all groups were required to learn a second set of responses to the same set of stimuli.

All groups that received preliminary training did better than the group that received no training. As predicted by Hypothesis 1, the group that overlearned the initial responses used cues from the initial responses to a greater extent than the group that did not overlearn the initial responses. The group with overlearning also demonstrated more transfer. However, this superiority was apparently not due to the greater use of initial responses as cues. As predicted by Hypothesis 2, there was no significant difference between the group that learned distinctive overt responses and the group that learned equivalent overt responses but distinctive perceptual responses, yet both of these groups showed more transfer than the group that had learned equivalent overt and equivalent perceptual responses.

References

ARNOULT, M. D. Stimulus predifferentiation: Some generalizations and hypotheses. *Psychol. Bull.*, 1957, **54**, 339–350.

CANTOR, G. N. Effects of three types of pretraining on discrimination learning in preschool children. *J. exp. Psychol.*, 1955, **49**, 339–342.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy.* New York: McGraw-Hill, 1950.

GAGNE, R. M., & BAKER, K. E. Stimulus pre-differentiation as a factor in transfer of training. *J. exp. Psychol.*, 1950, **40**, 439–451.

GOSS, A. E. Transfer as a function of type and amount of preliminary experience with the task stimuli. *J. exp. Psychol.*, 1953, **46**, 419–428.

HAKE, H. W., & ERIKSEN, C. W. Effect of number of permissible response categories on learning of a constant number of visual stimuli. *J. exp. Psychol.*, 1955, **50**, 161–167.

JAMES, W. *The principles of psychology.* New York: Dover Publications, 1950.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulis. *J. exp. Psychol.*, 1949, **39**, 770–784.

MILLER, N. E., & DOLLARD, J. *Social learning and imitation.* New Haven: Yale Univer. Press, 1941.

QUENOUILLE, M. H. The analysis of covariance and non-orthogonal comparisons. *Biometrics*, 1948, **4**, 240–246.

ROBINSON, J. S. The effect of learning verbal labels for stimuli on their later discrimination. *J. exp. Psychol.*, 1955, **49**, 112–114.

(Received May 26, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

CHANGES IN PERFORMANCE AS A FUNCTION OF SHIFTS IN THE MAGNITUDE OF REINFORCEMENT[1]

GEORGE COLLIER AND MELVIN H. MARX

University of Missouri

Shifts in magnitude of reinforcement result in changes in performance. The course of the change and the final limit of change are two questions of major interest.

Both overshooting of the expected asymptote of performance when the magnitude is increased and undershooting when the magnitude is decreased have been reported (Crespi, 1942, 1944; Spence, 1956; Zeaman, 1949). Both gradual and abrupt changes in performance following a change in magnitude of reinforcement have been observed (Crespi, 1942, 1944; Dufort & Kimble, 1956; Pereboom, 1957a, 1957b; Spence, 1956). In these previous experiments on shifts in reinforcement, *S*'s experience with the first magnitude of reinforcement was gained concurrently with the acquisition of the instrumental response; the shift in reinforcement occurred at some point after asymptote had presumably been reached (Spence, 1956). In such a procedure the acquisition of the instrumental response and the experience with the reinforcement value are confounded. The present experiment examines the effect of shifts in magnitude of reinforcement when the experience of the first magnitude of reinforcement and the acquisition of the instrumental response occur independently.

[1] This investigation was supported by a research grant, M-817, from the National Institute of Mental Health, of the National Institutes of Health, Bethesda, Maryland. A brief report on the data was made at the meetings of the Midwestern Psychological Association, Detroit, 1958. The authors wish to thank Gopaul Sharma for his very careful collection of the data.

METHOD

Apparatus.—Four Skinner boxes delivering liquid reinforcements were used. The magazines were 12-in. diameter aluminum (free-machining) plates, $\frac{3}{8}$ in. thick, with seventy-two $\frac{1}{16}$-in. diameter cups drilled around the periphery. The maximum volume of each cup was .4 ml. A second plate on top of the first covered the cups and prevented evaporation. The plate was advanced by a Geneva movement driven by a synchronous motor. The operation of the magazine was barely audible. A Gerbrands bar activated by a force of approximately 22 g. was used. The magazine was located below the rat cage, which was $8 \times 9\frac{3}{4} \times 7$-in. cage of the same sort used to house the animal. The whole assembly was located in a soundproofed, light-tight, ventilated ice-box hull. The box was dimly illuminated.

Subjects.—Sixty female albino rats of the Sprague-Dawley strain acquired from the Holtzman Company were placed on a 23-hr. privation routine a week prior to the beginning of the experiment. The hour of feeding followed the experimental session by 1 hr. Losses through experimental error, failure of *S*s to begin taking all rewards within four days, sickness, and random discards to equalize the number in each group reduced the number of *S*s to 45.

Procedure.—Three groups of 15 *S*s were magazine trained on 4%, 11.3%, and 32% sucrose solutions, respectively, for eight days

$$\left(\%\ \text{conc.} = \frac{\text{wt. solute}}{\text{wt. solute} + \text{wt. solvent}} \times 100\right)$$

The magazine operated automatically on an aperiodic schedule with a geometric average of 1 min. The times between presentations were $\frac{1}{4}$, 1, 2, and 4 min. Twenty rewards were offered in the 20-min. session. All groups were then B-P (bar-press) trained on 11.3% sucrose solution, directly on a 1-min. periodic reinforcement schedule for ten days. The use of a test cage of the same kind as *S*'s home cage and of an aperiodic schedule during magazine training produces

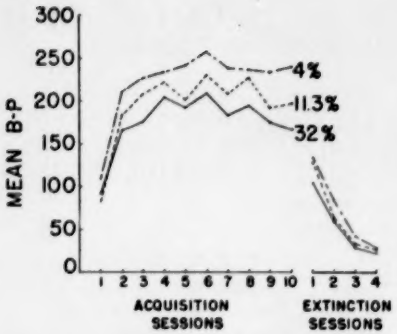

FIG. 1. Mean number of bar presses per 20-min. session for *Ss* bar-press trained on 11.3% and magazine trained on 4%, 11.3%, or 32% sucrose solutions.

very rapid conditioning of the B-P. Four days of extinction followed.

Skinner boxes and positions in the home cage rack were counterbalanced over groups.

RESULTS

The number of reinforcements taken was determined for each rat during magazine training. An analysis of variance[2] of these numbers showed no sigificant difference among groups, $P > .05$ ($F = 1.18$; $df = 2$ and 42). The mean number of reinforcements taken was 151.1 out of the 160 offered.

The mean number of bar presses for each group for each session during acquisition and extinction is presented in Fig. 1. An analysis of variance for Groups, Trials, and Replications is presented in Table 1. The main effects of Groups and Trials were significant, while their interaction was not. Replications and their interactions failed to reach significance. A test of the homogeneity of variance among groups failed to reach significance, $.20 < P < .30$ ($\chi^2 = .379$; $df = 2$). Individual t tests between the 4% and 11.3% groups and between 11.3% and 32% groups did not reach significance. The average difference between the 4% and 11.3% groups was slightly greater, 28.9, than that between the 11.3% and 32% groups, 18.5.[3] The group trained on 4% and tested on 11.3% was superior to the group trained on 11.3% and tested on 11.3%, while the group trained on 32% and tested on 11.3% was inferior. There was no trend over the 10 sessions which would indicate that these differences were diminishing. A similar analysis of variance on the extinction data showed the same results, the main effects of groups and trials being significant while their interaction was not.

TABLE 1

ANALYSIS OF VARIANCE OF NUMBER OF BAR PRESSES

Source	*df*	*MS*	*F*
Between *Ss*	44		
Groups (G)	2	88,543	3.27*
Replications (R)	2	20,879	.80
G × R	4	32,674	1.25
Error_b	36	26,195	
Within *Ss*	405		
Trials (T)	9	66,971	37.95**
T × G	18	1,183	.67
T × R	18	2,464	1.40
T × G × R	36	1,587	.90
Error_w	324	1,765	
Total	449		

* $P \leq .05$.
** $P \leq .01$.

[2] The dropping of *Ss* obviously restricts the variance. The number dropped did not itself differ from group to group. It has been our experience that two to three contacts with the solutions is sufficient to lead to asymptotic respondings, whereas *Ss* that do not drink within the first two to three days almost never do.

[3] An examination of these differences by a medians test yielded probabilities between the 5 and 10% level for both differences, indicating that the failure of the t tests to reach significance is not due to skewness.

Discussion

The occurrence of abrupt changes in performance following shifts in the magnitude of reinforcement, as well as their bidirectional character, has led to the identification of the magnitude of reinforcement with motivational rather than habit factors (Dufort & Kimble, 1956; Hull, 1951; Seward, 1956; Spence, 1956). The failure of the limits of the course of change to conform with the asymptotic values obtained with *S*s trained on the new magnitudes has led to the introduction of additional emotional factors, dramatically called elation and depression (Crespi, 1942, 1944) or to the assumption that the effective reinforcement on any trial is determined by the previous history of magnitudes of reinforcement (Jones, 1952; Spence, 1956).

Three sorts of mechanisms have been invented to collect the properties of the magnitude of reinforcement: (*a*) emotions with drive properties resulting from the delayed gratification of the fractional consummatory response (Spence, 1956) or the frustration of expectations (Crespi, 1942, 1944); (*b*) motivational properties associated in an unspecified fashion with the magnitude of the fractional goal response, r_g, (Spence, 1956); and (*c*) variation of the effective reward as a function of expectation determined by the chain of preceding reinforcement magnitudes (Jones, 1952). In the first two mechanisms, which were introduced in part to account for abrupt shifts in performance via motivational changes, no provision is made to handle the effects of preceding reinforcement experience on present reinforcement effectiveness. Spence's (1956) association of K (incentive motivation) with the magnitude of r_g, which is classically conditioned, implies that K should change as a function of trials following a shift in reinforcement magnitude. His interpretations of his own data belie this implication in that K was introduced to account for the abrupt shifts in performance. Spence notes in several places the occurrence of such abrupt shifts in his and other data (e.g., Spence, 1956, pp. 130–133). This difficulty with K was long ago pointed out by Leeper (1952). Jones (1952), in an attempt to deal with this problem, has modified Hull's postulates to substitute expectation for habit. His formulation is based on the assumption that the effective value of a reward is related to both the present magnitude and the preceding sequence of rewards. He has formalized this assumption in a way such that the other relations generated by Hull's ${}_sH_R$ and ${}_sE_R$ concepts remain unchanged.

The results of the present experiment seem to offer difficulties for each of these mechanisms. Neither the fractional goal response nor emotional states produced by lack of immediate gratification can handle the obtained outcome. On the assumption that the magnitude of the r_g determines K, one would predict that the 32% group would have the r_g of greatest rate of responding, while the 4% group should show the lowest.[4] Further, using the notion of the classically conditioned r_g, one would expect all values to converge over trials to the level of performance of the 11.3% group. Similarly, the present data rule out any interpretation in terms of motivation produced by frustration or delayed gratification of the r_g since, again, both the direction and duration of the effect are wrong. An account in terms of elation or depression is not plausible since the duration of the effect is too long. It seems unlikely that one would be still celebrating or complaining 10 days and 200 reinforcements later. While Jones' account takes care of the problem of the effect of preceding rewards, it does not suggest the variables which govern the magnitude of the shift effect.

Thus it appears that the identification

[4] Preliminary results obtained in our laboratory indicate that this prediction is verified when a stimulus associated with different magnitudes of reinforcement for different groups during magazine training is used as a reinforcer. In this case, the greater the magnitude of reinforcement the greater the reinforcing power of the associated stimulus.

of magnitude of reinforcement directly with either the asymptote of habit strength or a motivational factor is incorrect.

Recent studies of the amount of reinforcement function have suggested that performance is a function of the stimulus dimensions of reinforcement (Collier, 1958; Collier & Siskel, 1959). In this case it would be expected that reinforcement would have the properties of a sensory scale. One of the variables known to influence sensory magnitude is the sequence of preceding judgments, as demonstrated, for example, by the adaptation level or the anchoring effects observed in the absolute judgment experiments. Similarly, the exposures to the different concentrations during magazine training would be expected to lead to a different "sweetness" for the 11.3% concentration for the three groups of rats. Thus, one could argue that the reinforcing effects of a given magnitude of reinforcement, that is, for example, its *sweetness, saltiness, bitterness*, etc., are a joint function of the classical psychophysical variables, among which are the sequence of preceding stimulations. Thus, if one assumes that reinforcement has psychological dimensions, one can account for many of the effects which have been observed following shifts in magnitude of reinforcement, as well as the over- and undershooting of the expected asymptotes. This view would require, as do the Jones and Spence formulations, that the curves of the present experiment should eventually converge. It seems reasonable to assume that the effects of preceding stimulations diminish with successive stimulation. The present experiment shows that their effects are not transitory but leaves unanswered the questions as to their duration and to the variables which govern their duration.

These comments obviously do not explain the shift effects, since the "judgmental" effects are not themselves explained. They merely amount to a reclassification of the class of events to which the shift effects belong.

The present design avoids the problem of the interaction of the effects of shifts in reinforcement with changes in the performance of the instrumental response. There are a number of pieces of data in the literature which suggest that this may not be a single problem. For example, in the studies so far reported, those which have used one trial per day have shown abrupt shifts (e.g., Crespi, 1942; Spence, 1956, pp. 130–131; Zeaman, 1949) while those studies using more than one trial per day have shown relatively more gradual shifts (Dufort & Kimble, 1956; Pereboom: 1957a, 1957b). There is a suggestion in some preliminary data from our laboratory and some of Guttman's (1953) data that the gradualness is even more pronounced if the shifts occur within a session, where the more transient residual effects of the preshift experience are still present. On the other hand, Pereboom (1957a) has made a case for the differential effects of competing responses (exploration of the runway in his case) for *S*s trained under different magnitudes of reinforcement. Further, both Spence (1956) and Pereboom (1957a, 1957b) have pointed out roles which the asymptote of responding can play in determining the sorts of effects which occur following shifts.

This diversity of outcomes with respect to the course of change in performance following a shift in magnitude of reinforcement suggests that the course of change may be influenced by a number of variables not associated with the magnitude of rewards. The "effectiveness" of a reinforcer is a joint function of the present value of the stimulus characteristics and previous contacts with these or other values. The character of the course of shift is a function of the situation in which it is determined.

Summary

Three groups of rats were magazine-trained on either 4%, 11.3%, or 32% concentrations of sucrose solutions. All groups were then bar-press (B-P) trained on 11.3%. The 4% group was significantly superior to the 11.3% and the 32% significantly inferior for the 10 days of rewarded B-P and the four days of extinction.

The consequences of these results for the motivational and habit interpretations of reward were discussed. It was argued that the effects which have led to these interpretations can be subsumed under the phenomena observed in determinations of psychological scales.

REFERENCES

COLLIER, G. The interaction of factors governing the amount of reinforcement function. Paper read at Midwest. Psychol. Ass., Detroit, May, 1958.

COLLIER, G., & SISKEL, M., JR. Performance as a joint function of amount of reinforcement and inter-reinforcement interval. *J. exp. Psychol.*, 1959, **57**, 115–120.

CRESPI, L. P. Quantitative variation of incentive and performance in the white rat. *Amer. J. Psychol.*, 1942, **55**, 467–517.

CRESPI, L. P. Amount of reinforcement and level of performance. *Psychol. Rev.*, 1944, **51**, 341–357.

DUFORT, R. F., & KIMBLE, G. A. Changes in response strength with changes in the amount of reinforcement. *J. exp. Psychol.*, 1956, **51**, 185–191.

GUTTMAN, N. Operant conditioning, extinction, and periodic reinforcement in relation to concentration of sucrose used as a reinforcing agent. *J. exp. Psychol.*, 1953, **46**, 213–224.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JONES, M. B. Effect, change, and expectation of reward. *Psychol. Rev.*, 1952, **59**, 227–233.

LEEPER, R. Review of Hull's *Essentials of Behavior. Amer. J. Psychol.*, 1952, **65**, 478–491.

PEREBOOM, A. C. An analysis and revision of Hull's theorem 30. *J. exp. Psychol.*, 1957, **53**, 234–238. (a)

PEREBOOM, A. C. A note on the Crespi effect. *Psychol. Rev.*, 1957, **64**, 263–264. (b)

SEWARD, J. P. Drive, incentive, and reinforcement. *Psychol. Rev.*, 1956, **63**, 195–203.

SPENCE, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

ZEAMAN, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received May 26, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

GROWTH OF A MOTOR SKILL AS A FUNCTION OF DISTRIBUTION OF PRACTICE

JOHN M. DIGMAN

University of Hawaii

While the influence of massing of practice on performance of a motor skill can be readily demonstrated, studies concerned with the effects of this variable on the rate of growth and asymptote of habit strength have often seemed inconclusive. Theoretical predictions to the contrary, a number of studies (e.g., Archer, 1954; Bourne & Archer, 1956; Digman, 1956; Reynolds & Bilodeau, 1952) have failed to find clear-cut evidence for any permanent decrement such as Hull's (1951) ${}_{s}I_{R}$, or the inhibition variables postulated by Ammons (1947) and Kimble (1949). Denny, Frisbey, and Weaver (1955) found "several lines of evidence" favoring the concept of conditioned inhibition in motor skill learning. However, the fact that the decremental function observed by them extinguished under distributed practice suggests a concept somewhat different from the traditional ${}_{s}I_{R}$ function. A similar phenomenon was observed by Digman (1956), who favored an explanation in terms of transfer theory, although the extinction interpretation is admittedly more basic.

Bourne and Archer (1956), measuring total time on target and time continuously on target (duration of hits), proposed that only one type of inhibition exists. This is the familiar temporary decrement, I_R, which is presumed to dissipate rapidly with rest. (However, Janhke and Duncan [1956] have found evidence for traces of I_R after 24 hr. of rest.) Concerning I_R, Bourne and Archer (1956, p. 31) suggested that ". . . in addition to reducing performance . . . it interferes with the acquisition of skill and thereby reduces the actual level of learning of the habit . . . [especially that part associated with] smooth pursuit tracking skill."

A study by Adams (1952), which was indirectly concerned with ${}_{s}I_{R}$, suggested the possibility of a convergence of the performances of two practice groups, one with practice well distributed, the other with practice considerably massed. In this study there were five practice sessions on as many consecutive days. Whether the performance of the massed practice group would have risen to the level of the performance curve for the distributed practice group, had there been another day of practice under distributed practice conditions for both groups, is a question immediately suggested by an inspection of Adams' curves, particularly in view of the conclusions reached by Bourne and Archer. It is this question which prompted the study reported here.

METHOD

Subjects.—The *S*s were 41 men and women drawn from two classes in psychology conducted by the investigator. Approximately half of the *S*s were volunteers from an introductory course; the rest were students in a course in the psychology of learning, who participated in the experiment as an essential part of the course. In general, *S*s were assigned to experimental conditions in order of their appearance except for an attempt to equalize sexes in the groups. One *S* was assigned, out of order, to the distributed practice group, Group D. The *N* in Group D was 22 (11 men and 11 women), and the

N in Group M, the massed practice group, was 19 (10 men and 9 women).

Apparatus.—A pursuit rotor manufactured by the Ralph Gerbrands Co. was used. The target button was 12.5 mm. in diameter, imbedded in a disc 25.5 cm. in diameter. The target button revolved at the rate of 30 rpm in a clockwise direction and in an epicyclic fashion. A brass stylus, hinged 12.0 cm. from the target end, with a tip 5.5 mm. in diameter, completed a circuit with the scoring element of a Standard Electric clock, whenever the tip was in contact with the target button during a work period. Target-to-floor distance was 35 in. An electronic interval timer manufactured by the Photoswitch Co., with interval accuracy of approximately 99%, was used to obtain the desired work-rest intervals.

Procedure.—Six practice sessions were given to both groups over a two-week period on days Monday, Wednesday, and Friday. To eliminate bias associated with the longer rest from Friday to Monday, the first practice session was staggered, with S_1 beginning on Monday, S_2 on Wednesday, etc. Each practice session consisted of eighteen 30-sec. trials. For Group D there was a rest interval of 1.5 min. between trials throughout the experiment. For Group M this interval was 2 sec., which provided *E* the opportunity to read and reset the scoring clock. On Day 6 both groups were treated alike, with 1.5-min. intervals between trials. Following Trial 108 there were four "test trials."

Throughout the experiment an attempt was made to motivate *S*s to maximum effort by stressing the alleged importance of the research, by asking each *S* to "do his best," and by reporting *S*'s score to him at the conclusion of each trial.

At the final practice session, *S*s were told at the conclusion of Trial 108 that there was to be a test of ability consisting of four trials. The *S*s were again told of the importance of the experiment and urged to do their best.

Results

The over-all empirical results may be seen in Fig. 1. Something of the convergence effect obtained by Adams (1952) may be noticed, a convergence which seems more evident from a comparison of the terminal performances of both groups on succeeding days, and less so from a comparison of initial performances. At the beginning of Day 6, the difference in performance between the two groups is approximately equal to the initial difference for Days 4 and 5, and not much less than that observed for Days 2 and 3.

Performance for Group M on Day 6, when practice conditions were

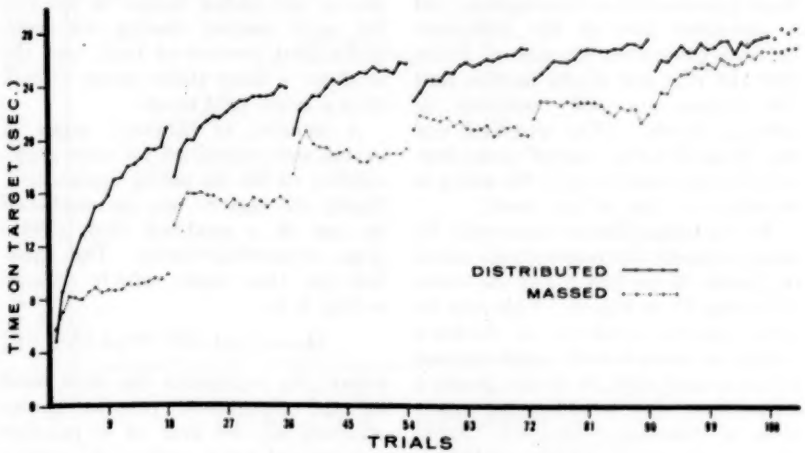

FIG. 1. Mean performance for distributed and massed practice groups.

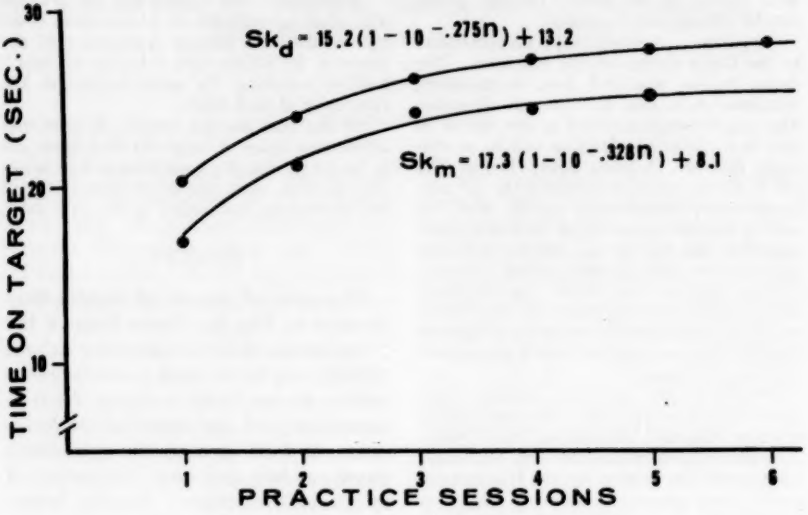

FIG. 2. Curves of skill growth for distributed and massed practice.

identical to those of Group D, does lend support to the original hypothesis that, had practice continued indefinitely, the performance curves for the two groups would have converged. However, the data suggest rather than demonstrate a convergence, and a one-sided test of the difference between groups for the sum of Trials 109–112 (the test trials) implies that the groups are still operating at different levels. (The test used was the Mann-Whitney signed ranks test, and the obtained z was 2.56, which is significant at the .01 risk level.)

An interesting resemblance may be noted between the performance curve for Group M on Day 6 and the curve for Group D on Day 4. This may be some indirect evidence of Archer's (1954) contention that under massed practice conditions *S*s do not practice as much as they would under conditions of distributed practice. Some of the implications of this resemblance are more fully considered below.

Parameters of growth in skill.—As an index of skill level (*Sk*) for Group D, the mean of the last two trials of each practice session was computed. This index was used for two reasons: the presence of warm-up decrement during the earlier stages of practice for each session, lasting for some unspecified amount of time, and the need for a more stable index of skill than a single trial mean.

A number of different types of curves were examined for their applicability to the six points determined. Easily the best fit was accomplished by use of a modified Hull (1951) type exponential curve. The equation for this curve, which appears in Fig. 2, is:

$$Sk_d = 15.2(1-10^{-.275n}) + 13.2 \quad [1]$$

where Sk_d represents the skill level for the distributed practice group attained at the end of a practice session, and n the number of practice sessions.

Because of the extensive amounts of work decrement usually present for Group M (except for Day 6), a different procedure was followed to determine skill level at the conclusion of each practice session under massed practice. It was first assumed that the amount of intersession forgetting was the same for both groups. (Indeed, there is evidence [Jahnke & Duncan, 1956] for little, if any, forgetting of rotary pursuit skill over a 2- or 3-day period.) A second assumption was that differences in performance at the beginning of any session reflected differences in skill level at that point. This difference, present at the beginning of a session, was then subtracted from the value for skill level for Group D at the conclusion of the previous practice session, and used as an index of skill level for Group M. Through the five points thus obtained, a modified Hull-type exponential curve was passed. The equation for the curve, which appears in Fig. 2, is:

$$Sk_m = 17.3(1-10^{-.328n}) + 8.1 \quad [2]$$

A measure of warm-up decrement, present at the beginning of each practice session, was computed by subtracting the obtained score for the first trial from the skill level attained at the end of the previous session. These values appear in Fig. 3, together with an extrapolation to zero practice. (Due to the manner of computing skill level for Group M, warm-up values computed in the above fashion would be the same for either practice condition.)

Shrinking the practice scale for Group M.—Since the performance of Group M on Day 6 bore such a resemblance to the performance of Group D on Day 4, it was tentatively concluded that at the beginning of Day 6, after 90 trials, Group M had received about as much practice as Group D had received at the beginning of Day 4, after 54 trials. Therefore, it was inferred that, under conditions of massed practice, at least for

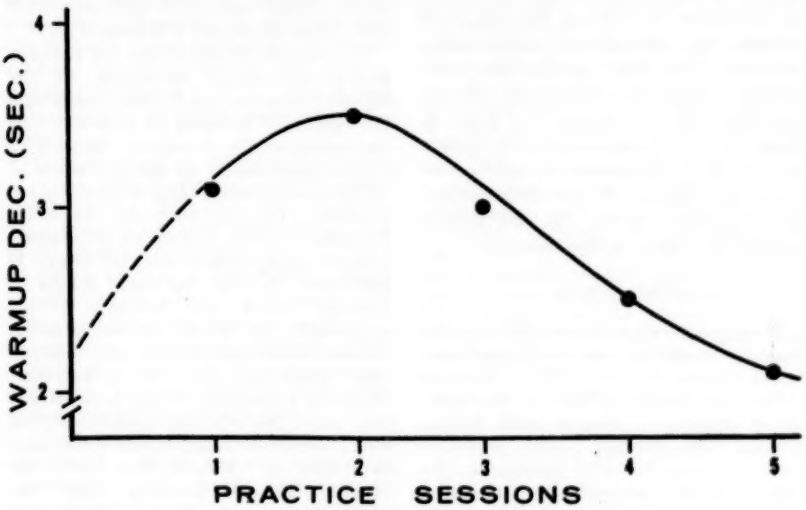

FIG. 3. Warm-up decrement as a function of practice level.

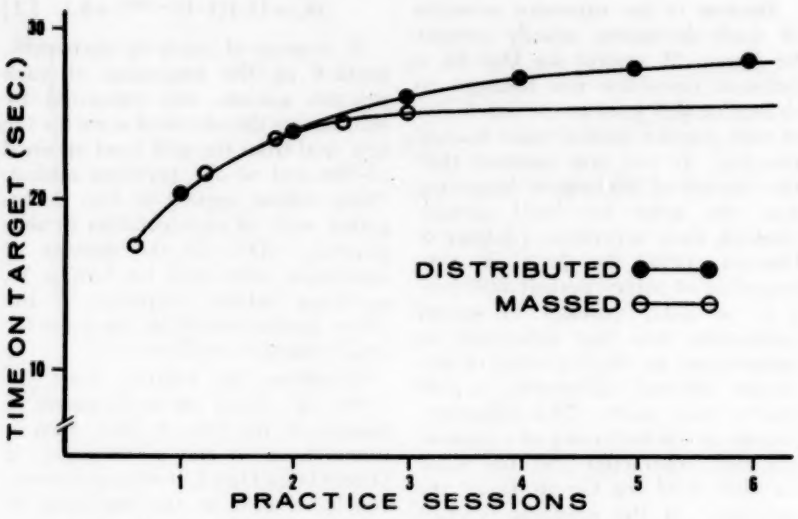

FIG. 4. Theoretical curves of skill growth (values for massed practice on shrunken practice scale).

the first three practice sessions, Group M received .60 as much practice per trial as Group D. The practice scale was accordingly shrunken to .60 of the scale for Group D, and skill values, as determined previously, plotted. The skill values for both groups, those for Group M on the shrunken scale, appear in Fig. 4. Here the curve and values for Group D are from Equation 1, while the curve for Group M has been drawn by inspection, using the asymptote and other values of Equation 2.

DISCUSSION

Two alternative interpretations of the data are available: the traditional two-factor theories of Hull (1951), Ammons (1947), and Kimble (1949); or the single-factor theory of Bourne and Archer (1956). The present study provides some evidence for the former in the failure of the performance curve for Group M to rise to the level of Group D on the last day of practice. The curve for Group M on the final day does, however, strongly suggest a convergence, and it may be argued that there was insufficient practice on the final day to develop those aspects of skill associated with "smooth pursuit tracking skill."

The results of the study are perhaps better understood according to the views of Bourne and Archer concerning the effects of massing of practice. As the curves in Fig. 4 suggest, massing of practice apparently set an asymptote of skill somewhat lower than did distributed practice. The fact that the values for Equation 2, when plotted on a shrunken practice scale, coincide with the Group D skill curve during the early stages of practice, bears out Archer's (1954) contention that under massed practice *Ss* learn less in part because they practice less. Beginning with Day 3, however, Group D apparently began to learn not only more per trial but also something which Group M failed to learn (at least until Day 6). This is very likely the "smooth pursuit tracking skill" for which Bourne and Archer (1956) have some evidence. The failure of the per-

formance curve for Group M to rise to the level of Group D may be attributed, as the curves in Fig. 4 suggest, to a considerable difference in practice in those parts of skill associated with high-level performance. Since the curves would imply that *Ss* in Group D began to learn these elements of skill perhaps as early as the beginning of Day 3, Group D may have had in this respect, by the end of the experiment, a fourfold advantage in practice over Group M.

While the computed skill values are unusually well described by the equations derived for them, they obviously fail to coincide at the point of zero practice. Extrapolation of the skill curve for Group M to the point of zero practice gives a value of 8.1. Actual performance for Group M on Trial 1 was 5.71. The difference between this and theoretical skill level, 2.4, is, by reference to the warm-up decrement curve, Fig. 3, a reasonable value. However, the skill curve for Group D, when extrapolated to zero, gives a value of 13.2. Actual performance was 4.83, which, even with the addition of 2.5 for warm-up decrement, gives a value considerably short of the skill level for zero practice derived from the equation. A possible interpretation of this state of affairs is that the equation is not applicable during the earliest part of practice, and that a different and more rapid growth of skill occurs under distributed practice at this stage.

The advantage of distributed practice over massed practice is, then, possibly threefold: (*a*) More learning occurs per trial under distributed practice, because more practice is received. (*b*) The growth of skill under distributed practice continues to a higher final level, very likely because of the acquisition of a type of response difficult to achieve under massed practice. (*c*) For some reason, early learning under distributed practice may be especially advantageous.

However, the repeated failure to find conclusive evidence for any permanent decrement, even under the severely contrasting conditions of this study, suggests that in some situations massing of practice, far from causing any permanent damage to skill, may actually be a very economical type of practice. While *Ss* in the present study who learned under massed practice conditions may have received only 60% as much practice per trial as *Ss* under distributed practice conditions, they spent only 9 min. in the experimental situation; the distributed practice *Ss* spent more than 30 min., most of which time was devoted to reading magazines or gazing out a window. An implication of the present study might be that some of this time could have been more profitably given over to practice, at least that portion of training after the beginning and before the acquisition of high-level skill.

Summary

The growth of a motor skill (rotary pursuit) was investigated, using two groups of *Ss* under contrasting degrees of distribution of practice. Both groups received eighteen 30-sec. trials per day for six days. For one group intertrial rests were 1.5 min. throughout the study. For the other group intertrial rests were 2 sec. for the first five days of practice and 1.5 min. on Day 6. A test of four trials, separated by 1.5 min., was given after Trial 108, under conditions assumed to be fairly optimal in respect to motivation and practice conditions.

Results indicate that the chief effect of massing of practice is to reduce the asymptote of skill, possibly by introducing conditions which make difficult the learning of the type of response necessary for high-level performance. It seems likely, however, that a switch in practice conditions, from massed to distributed practice, is accompanied by a corresponding shift in asymptote to that ordinarily associated with distributed practice. There was evidence that under massed practice *Ss* received considerably less practice per trial than did *Ss* under distributed practice. Possible evidence for a difference between "early" and "later" learning under distributed practice was tentatively offered.

References

Adams, J. A. Warm-up decrement in performance on the pursuit rotor. *Amer. J. Psychol.*, 1952, **55**, 404–414.

Ammons, R. B. Acquisition of motor skill: Quantitative analysis and theoretical formulation. *Psychol. Rev.*, 1947, **54**, 263–281.

ARCHER, E. J. Postrest performance in motor learning as a function of prerest distribution of practice. *J. exp. Psychol.*, 1954, **47**, 47–51.

BOURNE, L. E., & ARCHER, E. J. Time continuously on target as a function of distribution of practice. *J. exp. Psychol.*, 1956, **51**, 25–33.

DENNY, M. R., FRISBEY, N., & WEAVER, J., JR. Rotary pursuit performance nuder alternate conditions of distributed and massed practice. *J. exp. Psychol.*, 1955, **49**, 48–54.

DIGMAN, J. M. Performance under optimal practice conditions following three degrees of massing of early practice. *J. exp. Psychol.*, 1956, **52**, 189–193.

HULL, C. L. *Essentials of behavior.* New Haven: Yale Univer. Press, 1951.

JAHNKE, J. C., & DUNCAN, C. P. Reminiscence and forgetting in motor learning after extended rest intervals. *J. exp. Psychol.*, 1956, **52**, 273–282.

KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **39**, 15–23.

REYNOLDS, B., & BILODEAU, I. McD. Acquisition and retention of three psychomotor tasks as a function of distribution of practice during acquisition. *J. exp. Psychol.*, 1952, **44**, 19–26.

(Received June 4, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

RESPONSE DECREMENT, INDUCED BY STIMULUS CHANGE, AS A FUNCTION OF AMOUNT OF TRAINING [1]

DALBIR BINDRA AND JOHN F. SEELY

McGill University

Fink and Patton (1953) observed that the incidence of a learned drinking response decreased when the various visual, auditory, and tactual characteristics of the training situation were altered. They also found the response decrement to be proportional to the number of stimulus characteristics changed from the training to the test situation; the greater the stimulus change the greater the response decrement. The present investigation is concerned with exploring how the amount of training given prior to introducing a stimulus change affects the influence of that change on an instrumental response. Such an investigation was considered interesting because it could provide one direct test of Bindra's (1959; in press) formulation concerning the nature of stimulus-change induced response decrement. Furthermore, this formulation appeared to lead to predictions that are either opposed to those arising from the statements of Hull (1943, 1947) and of Guthrie (1935), Estes (1950) and Estes and Burke (1953), or do not necessarily follow from them.

According to Bindra's (1959; in press) formulation, stimulus change from the training to the test situation evokes some new, "irrelevant" responses, and it is the interference from these responses that leads to a decrement in the trained response. The new responses are considered to be reactions to the novelty provided by the changed (i.e., test) situation, and in test situations involving only a small stimulus change, the new responses are thought to be mostly of an "exploratory" nature. The degree of novelty or strangeness of the test situation is necessarily dependent upon the extent to which the subject has already become familiar with the training situation. Thus, for any given stimulus change, the greater the training in the training situation, the greater would be the novelty offered by the test situation, and hence the greater would be the interfering exploratory responses evoked during the tests. Assuming that the exploratory responses evoked by the training situation reach a stable level after a few training trials, the ratio of these responses to the exploratory responses evoked by the stimulus change would decrease with increased training in the training situation. It follows from these propositions that the greater the training given before introducing a stimulus change, the *greater* would be the relative response decrement brought about by that stimulus change.

$$\left(\text{Relative response decrement} = \frac{\text{training performance} - \text{test performance}}{\text{training performance}}\right)$$

[1] This study was supported by a grant (A.P. 27) from the National Research Council of Canada. The authors are grateful to S. Fillenbaum for his critical comments during the preparation of this paper.

Furthermore, according to this view, the more highly trained *Ss*, as compared to those trained less well, would, with continued reinforcement, require *longer* to reach the level of performance existing before the stimulus change, assuming that the greater the novelty presented by a stimulus change the longer will the interfering exploratory responses persist. To test these predictions, a stimulus change was introduced into a simple runway after groups of rats had received 15, 40, or 80 reinforced (thirst–water) training trials, and the reinforcement was continued during the tests until the performance of all groups reached an asymptote.

METHOD

Subjects.—The *Ss* were 36 male hooded rats about 100 days old. They were obtained from the colony maintained by the Royal Victoria Hospital in Montreal. A supply of Purina chow pellets was always available in their cages, but they were subjected to a 23-hr. water deprivation schedule throughout the experiment.

Apparatus.—A large piece of plywood, 45 × 12 in., served as the floor of the runway. The runway itself was 36 in. long and 6 in. high, and was made of ¼-in. plywood. The long sides of the runway were attached by removable bolts to a 12 × 6-in. plywood piece at each end; this construction made it possible to vary the width of the runway from 5 to 10 in. A starting box (8½ × 3 × 6 in.) was placed at one end of the runway and was separated from it by a guillotine gate. At the other end of the runway, a glass drinking tube was inserted through a hole in the end piece, 2 in. from the floor and equidistant from the two sides. The whole runway was covered with a removable ½-in. wire mesh top.

During training, the width of the runway was set at 5 in., and the floor was covered with sawdust. The stimulus change, introduced after the appropriate number of training trials, consisted of doubling the width of the runway to 10 in. and covering the floor with ½-in. wire mesh instead of sawdust.

Procedure.—On the first 8 days of the experiment, the (water-deprived) *Ss* were placed on a feeding stand to drink from water bottles for a period of about 40 min. All *Ss* became accustomed to drinking in this manner by the end of this period. On the following day each *S* was given two trials on the runway; each trial involved traversing the distance from the starting box to the drinking-tube end of the runway, and drinking water for about 10 sec. The time elapsing between the removal of the guillotine gate and the arrival of *S* at the other end of the runway was recorded with a stopwatch. Next day, *Ss* were given three trials each, and thereafter each *S* was given five daily trials. Immediately after the daily trials were over, each *S* was allowed 20 min. of free access to water on the feeding stand. Then it was returned to its cage.

The *Ss* were split into three equal groups of 12 *Ss* each. Groups I and III were equated by dividing 24 *Ss* between them on the basis of their performance on the first 15 trials; Group II was not equated with the others in this way, but its performance during training turned out to match remarkably well that of Group III. Group I was given a total of 15 training trials (none after grouping), Group II, 40, and Group III, 80 before the stimulus change was introduced. Following the stimulus change, *Ss* in each group were continued on the same schedule of five daily reinforced trials until the mean daily running time reached an asymptote.

RESULTS

Initial effects of stimulus change.—The mean running time for five daily trials (the initial five training trials, which were spread over two days, were considered as the trials for the first day of training) was calculated for each *S*, and the mean of these figures for each group was determined. These group mean running times are shown in Fig. 1. It is seen that there was no response decrement (i.e., increase in running time) following the stimulus change in the case of Group I; its mean running time on Day 4, the first test day, was *less* than its running time on Day 3, and its performance on Day 4 and on subsequent days (see the points on Day 8) was similar to the performance of the other groups. Groups II and III, on the other hand,

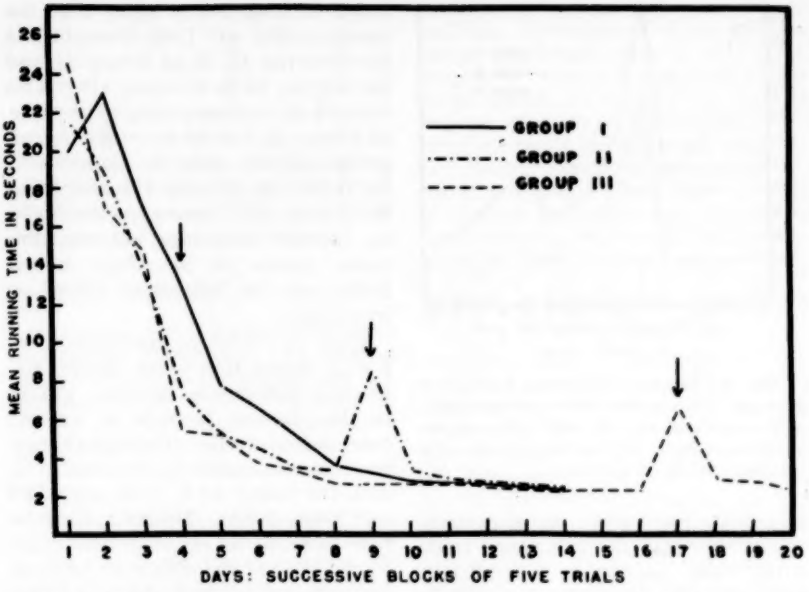

FIG. 1. Daily mean running times for the three groups of rats, which received 15, 40, or 80 trials before the stimulus change was introduced. The arrows indicate the first test day for the different groups.

showed a significant increase in the mean running time following the stimulus change; Group II's running time for Day 9 is significantly more than its running time on Day 8 ($P < .01$), and the mean running time for Group III on Day 17 is similarly different from the corresponding figure for Day 16 ($P < .01$). (All P's are based on t tests; the use of the appropriate nonparametric tests would not alter the reported findings.)

However, the stimulus change did lead to some response decrement even in the case of Group I, though this effect was overcome within two trials and does not show itself in the five-trial mean. This decrement is seen when the running times for individual trials just before and after the stimulus change are plotted, as in Fig. 2. Any possible "regression" in the strength of response from the last trial on one day to the first trial on the following day cannot account for the obtained decrement. No evidence for such regression could be found in any of the groups. For example, in the case of Group III, there was a median *decrease* in running time of 1.5 sec. from Trial 5 on Day 3 to Trial 1 on Day 4. Similarly, in the case of Group I, there was an improvement (median 1.3 sec.) in running time from the last trial of Day 2 to Trial 1 of Day 3.

Figure 2 shows that for Group I the first (test) trial after the stimulus change is as different in absolute terms from the running times immediately before the change as are the postchange running times of Groups II and III different from their prechange running times. When an

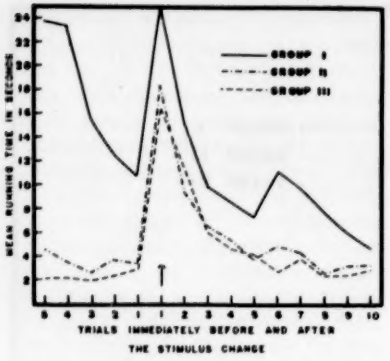

FIG. 2. Mean running time for Groups I, II, and III on each of the 5 trials preceding and on each of the 10 trials following the stimulus change. The arrow indicates the first test trial for all groups.

index of prechange running time (mean of the three fastest trials within the five trials immediately preceding the stimulus change) is compared with the running time for the first test trial, it is found to be significantly different ($P < .01$) from the latter for all the groups. This index of prechange running time had to be devised to serve as a stable measure of prechange running time in the case of Group I. The *S*s in all groups showed considerable trial-to-trial variation and a rapid decline in running time on Day 3 because training was still at an early stage. Since Day 3 was the last day of training for Group I, an index such as the one employed had to be devised in order to obtain a measure of prechange running time that was stable and that also incorporated the improvement in performance on Day 3. That this index does not grossly underestimate the running times on Day 3, thus adventitiously yielding the obtained decrement on the first test trial in the case of Group I, is suggested by the fact that this index for Day 3 was *higher* than the running time on Trial 1 on Day 4 for 7 of the 12 *S*s in Group II and for 9 of the 12 *S*s in Group III. This index was necessary only in the case of Group I, but it was used for all groups for the sake of consistency; by Trial 40 (Group II) and Trial 80 (Group III) there appeared to be no marked difference between this index (mean of the three fastest trials) and the daily mean (based on five trials).

Even a cursory examination of Fig. 2 shows that there are no significant differences between groups in the absolute increase in running time following the stimulus change, the mean increases for Groups I, II, and III being 13.4, 15.7, and 14.8 sec., respectively. However, the relative increase in running time, that is, the ratio of the difference between the index of prechange running time and test running time to the index of prechange running time, increased with increased training, the means being 1.7 for Group I, 5.8 for Group II, and 8.0 for Group III. The difference in this ratio between Group I and Group III is significant at the .01 level; and the difference between Groups I and II is significant at the .05 level, but the difference between Groups II and III is not significant.

Return to the prestimulus change level of performance.—In this experiment, reinforcement was continued during the test trials in order to determine any possible group differences in overcoming the effects of the stimulus change (see Fig. 2). The number of trials required by each *S* to reach, following the stimulus change, the level of its own index of the prechange running time was determined, and the group means calculated. These means were 4.3,

6.8, and 10.8 for Groups I, II, and III, respectively. The difference between Groups I and III is significant at beyond the .01 level, and the other differences at beyond the .05 level. These results show clearly that the greater the training given before the stimulus change was introduced the more trials it took *Ss* to overcome the response decrement.

Discussion

The results of the present experiment demonstrate clearly that, with increase in training prior to the introduction of a stimulus change, there is an increase in relative response decrement as well as in the number of reinforced trials required to re-reach the prestimulus-change level of performance. Both these results were predicted from, and are consistent with, Bindra's (1959; in press) formulation of the nature of stimulus-change induced response decrement, as described earlier. The present results are also relevant to the views of Hull (1943, 1947) and of Guthrie (1935), Estes (1950) and Estes and Burke (1953), and it might be profitable to consider how far our results are consistent with these views.

It is difficult to say what the Hullian prediction for this experiment would be because there is no unequivocal way of deciding how much of Hull's system to take into account. At a fairly simple level, his statements on "stimulus generalization" seem to apply to the present experiment. According to Hull, in the early stages of training, stimulus generalization is complete or nearly so, but, with increasing amounts of training, "a downward sloping gradient of primary stimulus generalization of reaction strength gradually and spontaneously develops" (Hull, 1947, p. 133). Hull considers generalized response strength also to be a function of response strength generated during training (Hull, 1943, p. 199), so that any training after a gradient has been established would result in progressively greater generalized response strength at each point on the gradient. Furthermore, if the stimulus change were made after a very large amount of training, though there should be an increase in the absolute level of response strength throughout the gradient, there would be little reason to expect any further change in response decrement; however, for any given amount of response decrement from the training to the test situation, there should be a decrease in *relative response decrement*, i.e.,

$$\frac{\text{training performance} - \text{test performance}}{\text{training performance}}$$

Applying these propositions to the present experiment, one should expect first an increase in relative response decrement, followed in this case where the absolute response decrements for the three groups are roughly the same, by a decrease in relative response decrement. However, the present results show a continuous increase in relative response decrement. As mentioned above, other predictions for the results of our investigation can also be derived by taking into account other aspects of Hull's principles, and it is possible that some of those predictions will be consistent with the present results. Thus, at present, no clear disproof of Hull's views is possible.

According to Guthrie (1935) and the formulations of Estes (1950) and Estes and Burke (1953), which are based on Guthrie's general viewpoint, the probability of occurrence of a trained response in the test situation is a function, roughly, of the number of stimulus elements in the training situation that are also present in the test situation. Inasmuch as the stimulus elements that become conditioned to the response with increased training are assumed to be randomly distributed within the total stimulus elements in the situation, the proportion of the conditioned stimulus elements operating after the stimulus change would seem to be the same at all levels of training. Therefore, one should expect that a given stimulus change would cause the same relative response

decrement irrespective of the amount of training given prior to the stimulus change. This prediction is also contrary to the findings of the present experiment. However, here again, by making certain assumptions about the relative weight of the new elements introduced by the stimulus change at different levels of training, a correct prediction can be derived from such a stimulus-elements model. It should be noted that, basically, Bindra's formulation is consistent with the view that the occurrence of a response depends upon the proportion of stimulus elements that are common to the training and the test situations. However, it further states that a change in stimulus elements produces response decrement by virtue of evoking exploratory responses that interfere with the trained response, and that the extent to which the exploratory responses are evoked is determined, at least partly, by the degree of familiarity developed with the training situation. Whether "degree of familiarity" can be incorporated into the formulations of Estes (1950) and Estes and Burke (1953) remains to be seen.

Summary

This study deals with the relation between the amount of instrumental training given prior to introducing a stimulus change and the influence of that change on the trained response.

A stimulus change was introduced into a simple runway after groups of rats had received 15, 40, or 80 reinforced (thirst–water) training trials, and the reinforcement was continued during the tests until all groups reached an asymptote. The results showed the relative response decrement brought about by the stimulus change to increase as a function of the number of training trials. It was also found that the greater the training given prior to introducing the stimulus change the more (reinforced) trials it took the animals to overcome the response decrement (i.e., to return to the prechange level of performance).

It was suggested that these results require further clarification of the formulations of Hull, Guthrie, and Estes, but are quite consistent with Bindra's view concerning the nature of stimulus-change induced response decrement. According to the latter view, the amount of response decrement depends upon the degree to which the novelty provided by the stimulus change evokes interfering "exploratory" responses; and the novelty of a given stimulus change is thought to increase with increased training under the conditions existing before the stimulus change is introduced.

References

Bindra, D. *Motivation, a systematic reinterpretation.* New York: Ronald Press, 1959.

Bindra, D. Stimulus change, reactions to novelty, and response decrement. *Psychol. Rev.*, in press.

Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.

Estes, W. K., & Burke, C. J. A theory of stimulus variability in learning. *Psychol. Rev.*, 1953, **60**, 276–286.

Fink, J. B., & Patton, R. M. Decrement of a learned drinking response accompanyiny changes in several stimulus characteristics. *J. comp. physiol. Psychol.*, 1953, **46**, 23–27.

Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.

Hull, C. L. *Principles of behavior.* New York: Appleton-Century-Crofts, 1943.

Hull, C. L. The problem of primary stimulus generalization. *Psychol. Rev.*, 1947, **54**, 120–134.

(Received June 6, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

STIMULUS GENERALIZATION AS A FUNCTION OF LEVEL OF MOTIVATION [1, 2]

DAVID R. THOMAS [3] AND RICHARD A. KING [4]

Duke University

Though it seems to be a generally accepted fact that the effect of an increase in motivation is to increase stimulus generalization, there are differing conceptions of the nature of this function. Within the Hullian framework (Hull, 1943), since generalization is a habit function, it would be expected that the effect of drive would be multiplicative—i.e., the generalization gradients of *S*s with similar training but with different drive levels during testing would be multiples of each other. Miller (1944), on the other hand, has suggested that such gradients would be parallel, i.e., additively related. The present study was designed to investigate generalization as a function of drive level, using different levels of food deprivation in pigeons, the key-pecking response, and the stimulus dimension of the visual spectrum.

EXPERIMENT I

Method

Subjects.—Thirty-seven white Carneau pigeons were used.

Apparatus.—An automatic key-pecking apparatus fitted with a Bausch and Lomb monochromator was used. This apparatus is described in detail by Hanson (1957).

Procedure.—The *S*s were individually caged and maintained ad libitum until their weight stabilized (approximately 7 days). They were then starved to 80% weight, magazine trained, and trained to peck at a key illuminated by a monochromatic light of 550 mμ. After 50 continuous reinforcements, 10 consecutive daily sessions of VI (variable interval) training were administered. The mean inter-reinforcement interval was approximately 1 min. Stimulus-on periods of 60 sec. were alternated with stimulus-off ("blackout") periods of 10 sec. Thirty stimulus-on periods constituted a daily training session.

The *S*s were divided into four groups matched on the basis of number of responses on Day 10 of VI training. The mean number of responses was 1279 for the group to be tested at 90% of ad libitum weight (Group 90, $N = 10$). For Group 80 ($N = 9$), the mean number of responses was 1339; for Group 70 ($N = 9$), $M = 1491$; and for Group 60 ($N = 9$), $M = 1373$. These differences are not significant ($F < 1$; 3, 33*df*).

For 7 days after the completion of VI training, *S*s were maintained at 80% weight, or were starved or fed to the level appropriate for their group. On Day 8 after the completion of VI, all *S*s were tested for stimulus generalization in extinction. Eleven different test stimuli were used (490 mμ to 610 mμ by 10 mμ intervals, with the omission of 500 mμ and 600 mμ). With the single exception that each test began with the CS, the 11 test stimuli were randomized within a series, and 12 different random series were presented to each *S*. This resulted in a schedule of 132 stimulus presentations. Each stimulus presentation was for 30 sec. and was followed by a 10-sec. blackout period, during which the number of key-pecks in the preceding period was recorded and the stimulus was changed.

After an interval of one day to permit spontaneous recovery, a second generalization test was administered.

[1] An earlier version of this paper was presented at the May, 1958, meetings of the Midwestern Psychological Association in Detroit. The research was carried out while the first author was a research fellow of the United States Public Health Service.

[2] The authors wish to express their appreciation to Norman Guttman for his invaluable contributions to all stages of this research. The study was supported by Grant M-1002 from the National Institute of Mental Health, U. S. Public Health Service and by the Research Council of Duke University.

[3] Now at Kent State University.

[4] Now at the University of North Carolina.

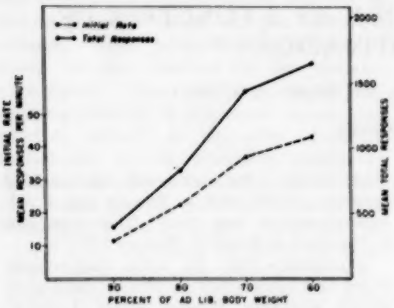

FIG. 1. Response strength as a function of body weight.

Results and Discussion

Although the relationship between response strength and hunger level is somewhat peripheral to the major purpose of this experiment, it may be of interest to examine this function before proceeding. Response rate at the start of the first generalization test (as determined from the first three presentations of all test stimuli) varies as an increasing monotonic function of deprivation level (Fig. 1). The differences among groups are significant ($F = 3.89$; 3, 33 df; $.01 < P < .025$).

Not only is the momentary rate a function of drive but also the total number of responses emitted during the 1½ hr. test varies in a similar fashion (see Fig. 1). These differences are also significant ($F = 3.46$; 3, 33 df; $.025 < P < .05$).

The mean generalization gradients obtained from the four drive groups are presented in Fig. 2. Though similar in general form, the gradients differ both in over-all height (level of responding) and in slope. The curves indicate that generalization, in terms of the total number of responses given to non-CS values of the stimulus, increases monotonically with increases in drive level.

If the obtained curves are extrapolated in the direction of both higher and lower wave lengths than those sampled, it appears that the extent of generalization also positively varies with drive. In fact it appears likely that *S*s at 60% weight would have responded with considerable strength to stimuli covering the entire range of the visual spectrum.

To make the slopes of the obtained generalization gradients of the four groups directly comparable (i.e., to adjust for differences in level of responding), the gradient of each *S* on generalization Tests 1 and 2 was transformed into percentages of total response output to form what may be called gradients of relative generalization. In Fig. 3 are presented the four mean gradients in percentage form. Because these gradients are not symmetrical, slope differences for left and right sides were analyzed separately. The Alexander trend tests are presented in Table 1. The probability levels for the individual F tests for differences between group slopes are presented in Table 2. These tests support the finding that the gradient of the 80% group is the steepest, that of the 60% group flattest, while the 70% and 90%

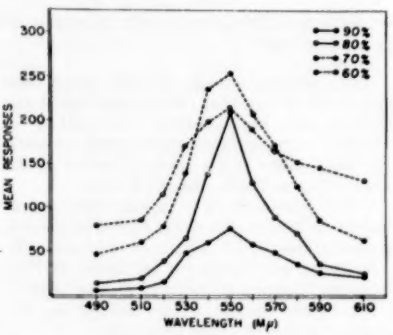

FIG. 2. Stimulus generalization gradients of the four experiment groups. (Exp. I.)

TABLE 1

ALEXANDER TREND TESTS OF RELATIVE GENERALIZATION GRADIENTS (EXP. I)

Source	*df*	≷550 mμ		≥550 mμ	
		MS	*F*	*MS*	*F*
Over-all slope	1	7585.10	265.12*	5116.95	68.71*
Over-all deviation from linearity	4	404.05	20.93*	149.53	9.25*
Between group slopes	3	203.26	7.10*	265.75	3.57***
Group deviation from linearity	12	39.39	2.02**	34.34	2.12**

* $P < .01$.
** $P < .025$.
*** $P < .05$.

group curves are intermediate in slope.

The generalization gradients obtained during spontaneous recovery were, on the whole, in agreement with those obtained on the first day of testing. The 60% group appeared to have the flattest gradient, while the other three gradients were quite similar to each other. Only the differences between the 60% curve and the 80% curve and the 60% and 70% curves (for CS and above) were significant ($P < .01$). None of the other slope differences between groups reached significance.

The finding of a relationship between response strength and drive level suggests the possibility of a relationship between response strength and generalization slope. To check this possibility, the three highest responding *Ss* (on Test 1) in each of the five[4] groups were pooled to form a high responding sample, while the three lowest responders in these groups were similarly combined to form a low responding sample. Though the high group responded five times as frequently as the low one, the relative generalization gradients of the two groups were quite similar (Fig. 4). There is little evidence in the present experimental situation for the hypothesis that relative generalization slope is significantly related to response strength. The effects on generalization slope here attributed to manipulations in drive may therefore not be ascribable simply to changes in response strength.

TABLE 2

SIGNIFICANCE LEVELS OF INDIVIDUAL GROUP SLOPE COMPARISONS (EXP. I)

	≷550 mμ			≥550 mμ		
	80%	70%	60%	80%	70%	60%
90%	<.05	N.S.	<.025	N.S.	N.S.	N.S.
80%		<.025	<.001		N.S.	<.005
70%			<.05			N.S.

Note.—N.S. = $P > .05$.

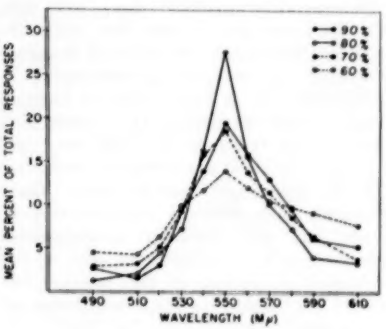

FIG. 3. Stimulus generalization gradients of the four experimental groups, in terms of percentage of total responses. (Exp. I.)

[4] Includes a group of *Ss* run in Exp. II.

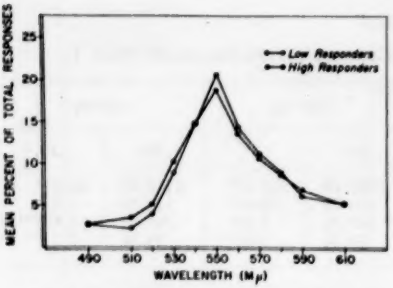

FIG. 4. Stimulus generalization gradients of high responding and low responding groups of *Ss*. (Exp. I.)

In a recent experiment similar in conception to the present one by Jenkins, Pascal, and Walker (1958), it was also found that generalization varies positively with motivation level. However, these investigators attributed the observed effect to the intervening variable of response strength. They present evidence that in the experimental situation which they employed, generalization slope varies reliably with response strength. They conclude that generalization slope varies with response strength, and with drive only indirectly.

Jenkins et al., employed a somewhat different training and testing procedure from the present one, and what is probably more important, they used a different stimulus dimension, circle size. It has been shown here that with the spectral dimension the effect of response strength on relative generalization is negligible. It is reasonable to suggest that even in the Jenkins et al. study, differences in response strength may not account for all of the variance attributable to the drive manipulation. Only research with different stimulus dimensions but with identical experimental procedures will indicate the role of the stimulus dimension in the relation between generalization slope and response strength.

EXPERIMENT II

Hull's multiplicative hypothesis would require the gradients of relative generalization (as here defined) to be the same for all drive groups. Miller's additive hypothesis would predict increasingly broader relative generalization gradients with increases in drive level. The results of Exp. I are clearly not in accord with either of these predictions. While the data are more clearly contradictory to the multiplicative hypothesis, the relatively flat gradient of the 90% group is inconsistent with both of the above conceptions.

Another explanation of the present findings consistent with Hullian theory involves the notion of the drive stimulus and its generalization. In the present study the steepest gradient was obtained with a group of *Ss* whose drive level was the same in testing as in training. The flattest gradient was obtained with the group which underwent the greatest drive change between training and testing. It is thus possible to hypothesize a flattening of the wave-length gradient as a function of drive stimulus changes. Such a notion would imply that when interoceptive stimulation is altered, exteroceptive stimulus changes become less important.

To test this possibility Exp. II was performed. This involved the training and testing of a group of *Ss* at 70% of ad libitum body weight. The generalization gradient of this group may be compared with those of two groups in Exp. I, the group trained at 80% and tested at 70% (80 → 70), and the group trained at 80% and tested at 80% (80 → 80). If the 70 → 70 group produces a gradient similar to that of the 80 → 80 group, the differences between groups in Exp. I may be attributed to a change in the drive stimulus from training to testing sessions. If the generalization gradient of the 70 → 70 group resembles that of the 80 → 70 group, the proposed role of drive stimuli in this situation would be open to serious question. If the gradient of the 70 → 70 group is intermediate between those of the 80 → 80 and the 80 → 70 groups, it would suggest that both drive stimuli and other characteristics of drive level have a role in determining relative generalization slope.

TABLE 3

ALEXANDER TREND TESTS FOR RELATIVE GENERALIZATION GRADIENTS OF GROUPS 80, 80 AND 70, 70 (EXP. II)

Source	*df*	≤550 mμ		≥550 mμ	
		MS	*F*	*MS*	*F*
Over-all slope	1	4921.23	130.02*	3559.61	59.53*
Over-all deviation from linearity	4	234.82	9.70**	138.03	9.10**
Between group slopes	1	243.83	6.44**	312.67	5.23***
Group deviation from linearity	4	48.99	2.02	76.09	5.02*

* $P < .01$.
** $P < .025$.
*** $P < .05$.

Method

Subjects.—Nine white Carneau pigeons were used.

Apparatus.—The apparatus was the same as in Exp. I.

Procedure.—The procedure was identical to that used in Exp. I with the exception that all nine *S*s were trained at 70% weight, maintained for seven days after the completion of VI training, and tested for generalization at this weight level.

Results and Discussion

The generalization gradient for the 70 → 70 group is presented in Fig. 5 with the 80 → 70 and 80 → 80 gradients from Exp. I. The similarity between the 80 → 70 and 70 → 70 gradients in Fig. 5 is striking. There is no significant difference between the slopes of these two curves. Table 3 presents the results of the analysis of the slopes of the 80 → 80 and 70 → 70 gradients. It is clear that these two curves differ significantly.

The results of Exp. II suggest that the slope of the generalization gradient varies with the drive level of *S* at the time of the test, and is independent of the drive level during training. Drive stimulus changes appear to have little or no effect on stimulus-generalization slope.

In conclusion, it was found that the amount of stimulus generalization in terms of the number of responses given to non-CS test stimuli varies as a monotonic increasing function of level of motivation. In addition, the extent of generalization, i.e., the range of stimuli which will evoke an appreciable response, seems to vary positively with drive.

Looked at in another way, however, the effect of drive is nonmonotonic. As hunger increases, the *proportion* of total responses given to non-CS values first decreases, then increases, with an apparent minimum at 80% of ad libitum weight.

The results of the present experiment appear to call into question any position which views response strength as a single entity to be considered independent of its various components. Perhaps habit and drive do not merely multiply with each other, but rather interact in

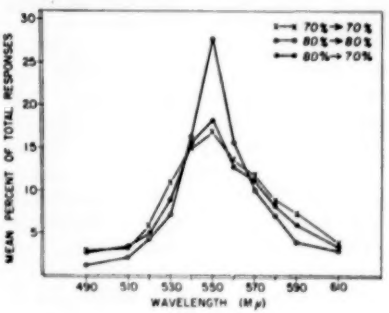

FIG. 5. Stimulus generalization gradients of 70 → 70, 80 → 70, and 80 → 80 groups. (Exp. II.)

some way. Indeed, it may be that different habits are learned and/or performed at different drive levels (Bruner, Matter, & Papanek, 1955; Cotton, 1953). In any case, it appears reasonably certain that as far as stimulus generalization is concerned, an *S* with low drive and high habit strength performs in a manner predictably different from an *S* with a higher drive but a weaker habit.[6]

Summary

The present investigation employed the key-pecking response in the pigeon to study stimulus generalization as a function of level of hunger. Four groups of pigeons were trained at 80% of ad lib. body weight to peck at a monochromatic stimulus, and then were tested at 90%, 80%, 70%, and 60% weight levels. It was found that in terms of absolute number of responses to stimuli other than the CS, or in terms of the range of stimuli to which *S* will respond, generalization increases as a monotonic function of deprivation. However, in terms of number of responses to non-CS stimuli relative to total responses, generalization first decreases, then increases, with a minimum at 80%.

Over-all response strength (both in terms of initial rate during generalization testing and total responses during the 1½ hr. test) appeared to be a negatively accelerated increasing function of weight loss. However, a comparison of the generalization gradients of high responding and low responding *S*s indicated that the effects of deprivation on generalization cannot be accounted for on the basis of changes in response strength alone.

In a second experiment to check on the role of the drive stimulus, it was shown that the slope of the generalization gradient varies with the drive level at the time of the test, and is independent of the drive level during training.

These results appear to call into question any position which views response strength as a single entity, to be considered independent of its associational and motivational components.

[6] A theoretical treatment of this problem which is in some respects consistent with the present findings has recently been offered by Estes (1958).

References

Bruner, J. S., Matter, J., & Papanek, M. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, **62**, 1–10.

Cotton, J. W. Running time as a function of amount of food deprivation. *J. exp. Psychol.*, 1953, **46**, 188–198.

Estes, W. K. Stimulus-response theory of drive. In *Nebraska Symposium on Motivation*. Lincoln: Univer. of Nebraska Press, 1958.

Hanson, H. M. Discrimination training effect on stimulus generalization gradient for spectrum stimuli. *Science*, 1957, **125**, 888–889.

Hull, C. L. *Principles of behavior*. New York: Appleton-Century-Crofts, 1943.

Jenkins, W. O., Pascal, G. R., & Walker, R. W. Deprivation and generalization. *J. exp. Psychol.*, 1958, **56**, 274–77.

Miller, N. E. "Experimental studies of conflict." In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald Press, 1944. Chapter 14.

(Received June 16, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

THE RELATION BETWEEN PHYSIOLOGICAL MEASURES OF EMOTIONALITY AND PERFORMANCE IN EYELID CONDITIONING [1]

W. N. RUNQUIST [2] AND L. E. ROSS

State University of Iowa

The basic theoretical framework within which recent studies of classical defense conditioning have been conducted assumes that drive level (*D*) and associative strength (*H*) combine multiplicatively to determine excitatory potential (*E*), which in turn determines the level of performance. It has further been assumed that drive level (*D*) in these studies is a function of a persisting emotional response (r_e) which is aroused by noxious stimulation and which, according to some neurophysiological evidence (Lindsley, 1951), may have a facilitative effect on cortical mechanisms. Since this mechanism (r_e) is assumed to have the properties of a response, its magnitude should be a function of those experimental variables which determine the strength of overt responses to noxious stimulation. This reasoning has led to the hypothesis that drive level (*D*) is a function of (*a*) the intensity of the noxious UCS used in conditioning, and (*b*) the score on a questionnaire designed to measure individual differences in emotionality, the A-Scale (Taylor, 1953). Confirmations of predictions based on this hypothesis and the other parts of the theory of conditioning given above have provided a good deal of support for the theory. This evidence has been extensively summarized by Spence (1956, 1958) and Spence and Ross (1957).

In addition to the use of the A-Scale, another method of selecting *S*s differing in emotionality is suggested by the theory. There are a number of so-called physiological responses such as the GSR, pulse rate, respiration rate, skin temperature, etc., which frequently have been used to define emotionality. The hypothesis may thus be formulated that the magnitude of these responses is some increasing function of the intensity of the hypothetical emotional response (r_e). If drive level is determined by the intensity of r_e, *S*s who give large physiological responses to noxious stimulation should have a higher level of drive (*D*) than *S*s who give small responses to this stimulation, and hence should exhibit a higher level of performance in eyelid conditioning.

In a preliminary study (Runquist & Ross, 1958) pulse rate changes and GSR responses to a weak air puff were measured, and *S*s were classified as emotional or nonemotional on the basis of these measures. While the emotional *S*s showed a significantly higher level of performance in eyelid conditioning than the nonemotional *S*s, certain difficulties involved in the measurement of the physiological responses made a repetition advisable under conditions which would give more reliable measures.

This study was designed to deter-

[1] This study, conducted under the general direction of K. W. Spence, is part of a project concerned with the influence of motivation on performance in learning under contract N9 onr-93802, Project NR 154–107 between the State University of Iowa and the Office of Naval Research.

[2] Now at Hobart College, Geneva, New York.

mine individual differences in emotional responsiveness to weak noxious stimulation, and to ascertain the relation of this level of responsiveness to level of performance in eyelid conditioning.

METHOD

The experiment consisted of two parts, both conducted during the same session. The first part was the testing period, the second part the conditioning period.

Test period.—Measures of each *S*'s pulse rate change and palmar skin conductance were taken in response to air puffs of medium intensity (.6 psi) delivered to the right eye at roughly 1-min. intervals. No warning signal was given. No measurements were made for the first five trials, since most *S*s gave a large skeletal reflex which interfered with accurate recording of the physiological responses. This could have been avoided by using a warning signal; however, most *S*s showed a response to any kind of warning signal which differentially confounded the magnitude of response to the air puff. With the equipment available simultaneous recording of both pulse rate and skin conductance was not possible. Therefore, pulse rate was recorded on the second five air puff trials and skin conductance on the following five. The measures obtained are thus confounded with differential adaptation; however, the effects of such adaptation would be to reduce individual differences and lessen the likelihood of obtaining the predicted results. Pulse rate and skin conductance were measured for a period of 15 sec., starting 5 sec. before the air puff.

Apparatus.—During the test period *S* was seated in the eyelid conditioning room. In addition to the regular equipment for presenting the UCS and recording eyeblinks (Spence & Taylor, 1951), one standard EKG electrode was attached to each forearm, and a pair of GSR electrodes consisting of lucite cups ¾ in. in diameter with a zinc plate at the bottom were placed on the back and palm of the right hand. The cups were filled with a cellulose sponge saturated with 10% zinc sulfate electrode paste.

Skin conductance was measured by the exosomatic method. The amplifier was a modified version of a simple bridge circuit described by Haggard and Gerbrands (1947). A measure of base level conductance was available on each trial. The EKG leads led to a standard EKG amplifier, the output of which led to a Brush ink writing oscillograph. Each pulse beat resulted in a vertical mark on a continuously moving paper.

Conditioning.—Immediately following the test period, eyelid conditioning was begun. Each *S* received 80 conditioning trials preceded by three trials with the CS alone. The CS was the increase in brightness of a 6-cm. circular disc. The UCS for conditioning was a 1-lb. air puff delivered to *S*'s right eye. The CS was 550 msec. in duration with the UCS occurring 500 msec. after its onset. The duration of the air puff was limited to 50 msec. by a 60-cycle 110-v. AC-operated solenoid valve controlled by an electronic timer. The intertrial interval was either 15, 20, or 25 sec. in a prearranged order. A verbal ready signal preceded each trial by from 2 to 4 sec. The *S*s were instructed to blink and fixate the CS upon presentation of the ready signal. A CR was recorded when the record showed a deflection of 1 mm. in the interval 200-500 msec. following the onset of the CS.

Subjects.—The *S*s were 90 volunteers from introductory psychology courses. An additional 32 *S*s were run but were discarded. Apparatus difficulties were responsible for 20, eight were classified as voluntary responders (Spence & Taylor, 1951), and four made two initial CR's to the light. All *S*s had previously taken the A-Scale.

Measurement of emotionality.—For each test trial, base level pulse rate was taken as the speed of the last four beats preceding the air puff, and skin conductance was measured immediately preceding the air puff. The response recorded was the fastest four beats and the highest conductance in the 10-sec. period following the air puff.

For the purposes of the main analysis, one response measure in each system was chosen: median pulse rate change, and change in base level conductance over the five test trials. The first measure was suggested by the previous study (Runquist & Ross, 1958) and the second by Schlosberg (1954).

RESULTS

Heart rate responses ranged from 44 beats/min. to minus 9 beats/min. with a mean change of 12.0 beats/min., while base level changes ranged from 30,000 ohms to minus 12,000 ohms with a mean change of 5800 ohms.

Two different methods of selecting emotional and nonemotional *S*s were employed. One involved converting

the median pulse rate change and change in base level conductance to z scores and then summing these to obtain an emotionality index. Using this method, the 20 *Ss* with the highest scores and 20 *Ss* with the lowest scores were compared on performance in conditioning. Curves of the acquisition of the CR in terms of percentage CR's in each block of ten trials are shown as the two solid curves of Fig. 1. The emotional *Ss* it will be seen, exhibit a consistently higher performance level over the course of conditioning. A Mann-Whitney test on the total number of CR's over the last 40 trials gave a z value of 1.56 which has a one-tail probability of .06. These results were obtained with an equal weighting of the two measures. Undoubtedly better prediction could be obtained by multiple regression analysis.

The second method of selecting *Ss* was suggested by some recent work on autonomic responses. Lacey (1956) has advanced the hypothesis that emotionality is response specific for each individual. That is, some people respond to noxious stimulation with an increase in pulse rate, others with an increase in skin conductance, others with respiratory changes, etc. This would lead to low correlations between the physiological measures and might lead to a low correlation between any single measure and performance in conditioning. The correlation between the two physiological measures was low ($r = .17$) as were the correlations between these two measures and conditioning (the r for pulse rate change and conditioning was .12, and for base level conductance change and conditioning was .29).

With this hypothesis in mind, two groups were selected as follows: An *S* was classified as emotional if he was

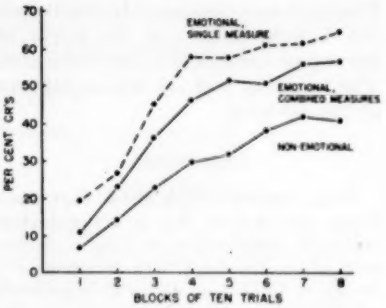

FIG. 1. Percentage of CR's in blocks of 10 trials for groups of *Ss* selected by combinations of physiological measures.

in the top 10% of the distribution on either pulse rate change (those showing the largest increase) *or* skin conductance base level change (those showing the largest increase). Eighteen *Ss* fit into this category. The nonemotional group was comprised of the same 20 *Ss* as the nonemotional group in the previous comparison. The curve of the acquisition of the CR is shown as the dotted line in Fig. 1. The emotional group is clearly superior to the nonemotional group in level of performance. A Mann-Whitney test of the total CR's in the last 40 trials gave a value of $z = 2.09$ which is significant at the .02 level. It appears, moreover, that selection of the extreme groups by this method results in greater differences in performance than the method of summing measures. The emotional *Ss* selected on the single measure criterion exhibited superior performance to that of the emotional *Ss* selected by the summed scores, although there is no way to evaluate this statistically since some of the same *Ss* appear in both emotional groups.

Since scores on the A-Scale were available for each *S* it was possible to make a determination of the relation between A-Scale score and physiologically defined emotionality. A

Pearson r was computed between each S's A-Scale score and the larger of his two physiological measure z scores. The resulting r of .22 was significant at the .05 level.

Discussion

These results which show that emotional *Ss* defined by a physiological index of emotionality exhibit a higher level of conditioning than nonemotional *Ss* are in agreement with the hypothesis which relates drive level (D) to the intensity of the emotional response (r_e). The results further support, although not as clearly, the hypothesis that there are individual differences not only in the level of emotionality, but in the particular response system which is activated.

In the preliminary study, it was found that the GSR response to the air puff was a fairly good predictor of conditioning performance, while in the present study it was not. The apparent contradiction is not surprising if one considers that each S received ten presentations of the air puff in this study before any GSR measures were taken. The effect of these trials would be to reduce any individual differences in the magnitude of the GSR response which were initially present.

The possibility should be considered that the changes in the physiological measures reflect skeletal startle responses to the air puff rather than emotional responsiveness. If such were the case a general drive theory would still predict the obtained conditioning, physiological index relationship since the magnitude of the startle reflex would also be assumed to reflect drive level (D).

Summary

Ninety *Ss* were given 15 air puff alone trials, on the last 10 of which pulse rate changes and skin conductance changes were measured. All *Ss* were then given 80 conditioning trials. Two groups of *Ss* were selected from the total sample and their conditioning performance compared. One group, the nonemotional group, were the 20 *Ss* with the lowest total combined z score on the two physiological measures. Emotional groups were defined in two ways: One emotional group consisted of the 20 *Ss* with the highest total z scores. Another emotional group was formed by taking the 10% *Ss* who showed the greatest pulse rate increase, and the 10% who showed the greatest skin conductance increase. The difference in conditioning performance between the nonemotional group and the emotional group selected by combining z scores approached significance, while the emotional group selected on single measures showed significantly higher performance than the nonemotional group. The results were interpreted as supporting a hypothesis which relates drive level (D) to emotional responsiveness.

References

Haggard, E. H., & Gerbrands, R. An apparatus for the measurement of continuous changes in palmar skin resistance. *J. exp. Psychol.*, 1947, **37**, 92–98.

Lacey, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123–164.

Lindsley, D. B. Emotion. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.

Runquist, W. N., & Ross, L. E. A preliminary investigation of the relation between heart rate change, GSR and level of performance in eyelid conditioning. *Proc. Ia. Acad. Sci.*, 1958, **65**, 398–404.

Schlosberg, H. Three dimensions of emotion. *Psychol. Rev.*, 1954, **61**, 81–88.

Spence, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.

Spence, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

Spence, K. W., & Ross, L. E. Experimental evidence on the relation between performance level in eyelid conditioning and anxiety (drive) level. *ONR Tech. Rep.* 5. State Univer. of Iowa, 1957.

Spence, K. W., & Taylor, J. A. Anxiety and strength of UCS as determiners of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183–188.

Taylor, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 284–290.

(Received June 20, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

SET TO RESPOND AND THE EFFECT OF INTERRUPTING SIGNALS UPON TRACKING PERFORMANCE [1]

STEPHEN GRIEW

Unit for Research on Employment of Older Workers, University of Bristol

It has now been demonstrated beyond much doubt that a delay will occur in making a response to a signal which occurs very shortly after another to which a response has to be made (Craik, 1948; Davis, 1957; Hick, 1948; Telford, 1931; Vince, 1949). Much of the work from which this conclusion has been drawn has been reviewed by Welford (1952) who also discusses the concept of "psychological refractoriness" which has arisen from it. This concept has been employed in the interpretation of a considerable range of high-speed performance (e.g., Hick, 1948; Hick & Bates, 1950; Welford, 1958) including tracking.

On the basis of the refractoriness hypothesis, one would predict that in a situation in which *S* is required to track continuously and at the same time to respond to irregular interrupting signals there will be some interference between the two sets of incoming signals, such that performance at both tasks will deteriorate. The first experiment to be reported below was designed to test this prediction. The aim of the second experiment was to clarify a suggestion arising from the first that the interference of signals from two sources is not the only determinant of performance decrement in tasks of this type.

[1] This research was part of a program financed by the Nuffield Foundation, London. The author is grateful to the Directors and Management of the Mond Nickel Company, Limited, for providing *Ss*, and to the *Ss* for their cooperation.

EXPERIMENT I

In an attempt to test the prediction outlined above, performance at a pursuit tracking task alone was compared with performance at a pursuit tracking task undertaken at the same time as responding to interrupting auditory signals. Also, efficiency at responding to interrupting auditory signals while maintaining continuous tracking was compared with efficiency at responding to auditory signals when tracking was not in progress. Interrupting signals were auditory in order to ensure that any interference which occurred would be centrally based.

Method

Apparatus.—The apparatus was a slightly modified version of that used by the author (Griew, 1958) in a study of age changes in pursuit tracking involving interrupted preview. A complex harmonic course, drawn on a band of paper, moved downward past a window placed perpendicularly to *S*'s line of sight. At the point at which it disappeared from view, a pen was placed. This pen could be moved across the paper by means of a handwheel. Preview of the course to be followed was given above the point at which matching took place. The *S* tracked the course by operating the handwheel with his left hand, and at the same time kept his right forefinger above a key placed conveniently on the right of the apparatus. This key was pressed immediately upon the occurrence of a signal from a buzzer operated by *E*. The signal from the buzzer was extinguished when the key was pressed. The tracking course took the form of 94 "swings" and did not repeat. It was the same at every tracking condition for all *Ss*. Interrupting auditory signals occurred at random time intervals apart, the intervals being the same for each

S, and ranging from 3 to 12 sec., with a mean of about 5 sec.

Subjects.—There were 20 *Ss*. They were industrial workers, randomly drawn from a population matched for occupation and "background." They were all right-handed. Their ages ranged from 21 to 50 yr. There were no age differences in efficiency of responding to auditory signals, and such age differences as were observable in performing the tracking task did not affect the issue being considered here; hence, the results of *Ss* of all ages are taken together.

Design.—Five experimental conditions were studied. These were:

Condition A: Responding to auditory signals alone. Here, *S* responded to a series of nine auditory signals given irregularly over the course of 45 sec.

Condition T_1: Performance at the pursuit tracking task alone at a course speed of 4.50 cm./sec. This gave a course duration of 64 (±1.20) sec.

Condition T_2: Performance at the pursuit tracking task alone at a course speed of 9.00 cm./sec. This gave a course duration of 32 (±0.60) sec.

Condition T_1A: Pursuit tracking at 4.50 cm./sec. and responding at the same time to 12 interrupting auditory signals.

Condition T_2A: Pursuit tracking at 9.00 cm./sec. and responding at the same time to seven interrupting auditory signals.

At each tracking condition, a 4.50 cm. preview of the course was given. Before performing for experimental purposes, all *Ss* performed at each experimental condition twice for practice purposes. The orders in which *Ss* attempted the five conditions were represented by the rows of a 5 × 5 latin square, replicated three times. Reaction times to auditory signals were recorded by the movement of a pen on the moving band of paper upon which the tracking course was drawn. Mean timing and amplitude errors in tracking were calculated in the manner described previously (Griew, 1958).

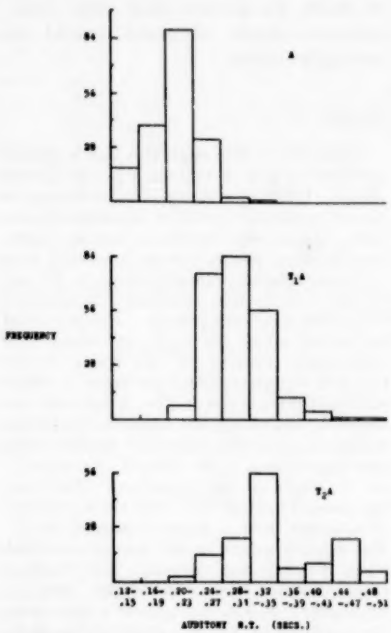

FIG. 1. Distributions of auditory RT's: Exp. I.

TABLE 1

TRACKING SCORES: EXP. I

Cond.	Mean Timing Error (Sec.)	Mean Amplitude Error (Cm.)
T_1	.01	.23
T_1A	.02 (P = ns)	.27 (P = .05)
T_2	−.05	.41
T_2A	−.05 (P = ns)	.54 (P = .002)

Results and Discussion

Mean timing and mean amplitude error scores in tracking are shown in Table 1. Table 2 shows mean auditory RT's with associated *SD*'s. The distributions of auditory RT's generating these means are shown in Fig. 1.

Summarizing these results, it may be said (*a*) that the introduction of interrupting auditory signals has had

TABLE 2

MEAN AUDITORY RT'S: EXP. I

Cond.	Mean Auditory RT (Sec.)	*SD*
A		.035
T_1A	(P = .05) .30 (P = .05)	.043
T_1A	.35	.069

no effect upon timing errors in tracking, (*b*) that the introduction of interrupting auditory signals has resulted in increases of amplitude error in tracking at both speeds, with a substantially greater increase at the higher speed, (*c*) that efficiency of responding to auditory signals has decreased upon the introduction of tracking at the same time, and (*d*) that the decreased efficiency of responding to auditory signals while tracking is accompanied by different distributions of auditory RT's at Cond. T_1A and T_2A. In Cond. T_1A, the distribution is very similar to that in Cond. A, but shows a higher mean, while in Cond. T_2A both a higher mean again, a much higher *SD* and a slightly bimodal distribution are shown.

The scatters of individual auditory RT's about their means in Cond. A and T_1A are so similar as to suggest that individual responses to interrupting signals are all being slightly delayed in the second. The bimodal distribution in Cond. T_2A suggests that the additional increase in mean RT in this instance is due to a much greater delay in a proportion of the responses to interrupting signals.

It is probably true that in higher speed tracking, interrupting signals are more likely to occur at a time when signals from the track (i.e., changes of direction) are being dealt with, and that, in consequence, their responses are likely to be sometimes greatly delayed. These delays are probably reflected in the cluster of high RT's in the bimodal distribution of Cond. T_2A results. The absence of high RT's in Cond. T_1A suggests that responses to auditory signals are probably tending not to be delayed because of interference by signals from the track. The general increase in auditory RT's in this condition is difficult to explain without recourse to a concept of "set to respond" primarily to signals from the track, with its implication that a greater degree of attention would be given by *S*s to the track, with the result that responses to auditory signals would, in general, all be slightly delayed. This would also account for the general increase in auditory RT in Cond. T_2A which have not been delayed by signals from the track.

On the basis of the foregoing discussion, two predictions may be made. First, at the lower speed, tracking performance would be expected to deteriorate only very slightly. This is because, with the possibility of anticipation given by a preview of 1 sec., and with attention being directed mainly at the track, it is unlikely that interrupting signals would interfere with individual responses to signals from the track. A nonsignificant increase in mean timing error, and a small but significant increase in mean amplitude error are, in fact, observed at the lower speed (see Table 1).

Secondly, at the higher tracking speed, the probability of signals from the track arriving while interrupting signals are being dealt with is higher. If a signal from the track arrives while a response to an interrupting signal is being initiated, it is likely that the track signal will not be dealt with until its response is almost due. By the time the response is made, it may be that the signal will have disappeared from view, and errors in matching would not then be adjustable. This should result in a much higher mean amplitude error than otherwise would have been the case. The results given in Table 1 confirm this expectation. Further, gross errors in matching amplitude of course "swings" would be expected when interrupting signals had just occurred.

In order to test this last prediction, the amplitude error exhibited at each course "swing" was divided by the amplitude of the "swing," to give an indication of *S*'s accuracy in matching amplitudes. This was done for each *S*, at each course "swing" at each tracking

TABLE 3

THE COINCIDENCE OF INTERRUPTING SIGNALS AND GROSS ERROR IN MATCHING AMPLITUDE OF COURSE "SWINGS" ($N = 17$)

Cond.	Total "Swings"	Number of Gross Errors in Matching (i.e. $\frac{a}{A} > 1$)	Number of Gross Errors Coincident With Interrupting Signals
T_1	1598	25	—
T_1A	1598	21	3
T_2	1598	28	—
T_2A	1598	109	78

condition. The results of this analysis are shown in Table 3.

In this table, $\frac{a}{A}$ indicates the ratio of amplitude error to amplitude of course "swing." High values of this expression, which indicate gross inaccuracies in matching amplitudes, occur most frequently in Cond. T_2A, and a high proportion of them coincide with points at which interrupting signals occur.

EXPERIMENT II

Although, on the basis of the results of Exp. I, it seems unlikely that the small increase in amplitude error in lower speed tracking with interrupting signals is due to interference between signals from the two sources, no indication is readily available as to the actual determinant of this small performance decrement. The most likely explanation is that there exists a secondary "set to respond" to interrupting signals which results in the occasional fluctuation of attention from the course to be followed to the source of interrupting signals. To test the hypothesis that a secondary "set" exists, tracking performance alone was compared with tracking performance in which *Ss* were warned to expect interrupting signals but in which no interrupting signals were, in fact, given.

Method

Apparatus and subjects.—The apparatus was the same as that used in Exp. I. Sixteen *Ss* took part. They were all right-handed, and were drawn from a similar population as were *Ss* in Exp. I. Their ages ranged from 25 to 47 yr. Such age differences as existed did not affect the issue under consideration and so the results of all *Ss* are taken together.

Design.—Four experimental conditions were studied. They corresponded to Cond. T_1, T_2, T_1A, and T_2A of Exp. I. The first two were identical to T_1 and T_2, representing tracking alone, and will be similarly designated T_1 and T_2. The last two differed from T_1A and T_2A in Exp. I in that, although *Ss* were warned to expect occasional interrupting signals, only one occurred in each of the corresponding conditions in this experiment. This was given, in both conditions, 12 sec. from the start of trials, and was intended to convey to *Ss* the impression that interrupting signals were, in fact, a serious possibility. The conditions corresponding to T_1A and T_2A will be designated Cond. T_1A' and T_2A'. In both experiments, instructions given to *Ss* were identical in the corresponding experimental conditions.

Before performing for experimental purposes, two practice trials at each condition were given to each *S*. During practice trials at T_1A' and T_2A', four interrupting signals were given, again to reinforce the impression that they were a serious possibility. The orders in which *Ss* attempted the four conditions were represented by the rows of a 4×4 latin square, replicated three times. Two *Ss* were drawn from each of eight departments of the factory in which the study was undertaken, and *Ss* from the same department performed immediately following one another, in order to minimize discussion of the experiment.

Results and Discussion

Table 4 shows mean timing and mean amplitude errors at the various conditions.

This experiment is concerned basically with the effect on pursuit tracking performance of the possibility of interrupting signals. It is clear that this possibility does not affect timing errors. The increase in mean amplitude error at the lower course speed

TABLE 4

TRACKING SCORES: EXP. II

Cond.	Mean Timing Error (Sec.)		Mean Amplitude Error (Cm.)	
T_1	.04	(P = ns)	.25	(P = .05)
T_1A'	.02		.28	
T_2	−.05	(P = ns)	.43	(P = .05)
T_2A'	−.07		.47	

when interrupting signals are possible is similar to that reported in the first experiment when interrupting signals were, in fact, given. At the higher tracking speed, the increase due to the possibility of interrupting signals is only about one-third of that observed in the previous experiment to be associated with the inclusion of interrupting signals.

It seems fairly clear that the "set to respond" to interrupting signals is sufficient to result in a decrement in pursuit tracking performance, which, while being slight, is nevertheless significant.

SUMMARY

Two experiments are reported. The first is concerned with tracking and responding simultaneously to interrupting auditory signals. Deterioration in efficiency at both tasks is observed when they are undertaken simultaneously. The manner in which these results are generated is discussed, and they are explained in terms (*a*) of the interference of signals from the two sources at the higher tracking speed and (*b*) of a "set to respond" primarily to signals generated by the tracking course. A subsidiary "set to respond" to interrupting signals is suggested by a slight but significant decrement in tracking performance at the lower speed. Experiment II investigates this by comparing tracking alone with tracking with the possibility of interrupting signals but when none are in fact given. A slight but significant decrement in tracking suggests that a "set to respond" is responsible for a proportion of performance decrement in tasks where interrupting signals are given.

REFERENCES

CRAIK, K. J. W. Theory of the human operator in control systems. II. Man as an element in a control system. *Brit. J. Psychol.*, 1948, **38**, 142–148.

DAVIS, R. The human operator as a single channel information system. *Quart. J. exp. Psychol.*, 1957, **9**, 119–129.

GRIEW, S. Age changes and information loss in performance of a pursuit tracking task involving interrupted preview. *J. exp. Psychol.*, 1958, **55**, 486–489.

HICK, W. E. Discontinuous functioning of the human operator in pursuit tasks. *Quart. J. exp. Psychol.*, 1948, **1**, 36–57.

HICK, W. E., & BATES, J. A. V. *The human operator of control mechanisms.* London. Ministry of Supply, 1950.

TELFORD, C. W. Refractory phase of voluntary and associative responses. *J. exp. Psychol.*, 1931, **14**, 1–36.

VINCE, M. A. Rapid response sequences and the psychological refractory period. *Brit. J. Psychol.*, 1949, **40**, 23–40.

WELFORD, A. T. The "psychological refractory period" and the timing of high speed performance—a review and a theory. *Brit. J. Psychol.*, 1952, **43**, 2–19.

WELFORD, A. T. *Ageing and human skill.* London. Oxford Univer. Press, 1958.

(Received June 10, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

EFFECTS OF INVERSION OF THE VISUAL FIELD ON HUMAN MOTIONS[1]

WARREN RHULE AND KARL U. SMITH

University of Wisconsin

The effects of inversion of the visual field were first investigated systematically by Stratton (1896, 1897, 1899), who used himself as a subject. Stratton reported especially on the difficulties of writing during inversion of the visual field. Recently interest has been revived in the problem by the work of Kohler[2] (1953), who has claimed that some of the difficulties in earlier studies can be overcome by requiring *S*s to wear reversing or inverting optics for many weeks rather than for only a few days. Kohler's work adds to earlier objective observations of Brown (1928), Ewert (1930, 1936), and Wooster (1923) in suggesting that readjustment to perceptual reversal is progressive, and is dependent on prior motor readjustment. In addition, studies have been conducted by Foley (1940), Hess (1956), and Kohler (1953) on the monkey and on the chicken, which suggest that readjustment to inversion or displacement of the visual field may be more difficult for lower animals, particularly the chick, than it is for man.

The one general finding which has come out of the human studies is that *S* can adjust somewhat both perceptually and motor-wise either to inversion or to right–left reversal of the visual field. The studies of both Ewert (1930, 1936) and Kohler (1953) suggest that this adjustment is not general, but is specific to task situations—that visual, tactual, and kinesthetic perception becomes consistent only after practice in and during performance of particular manipulative acts with objects or situations. The visual field is reorganized in relation to specific situations and performance in these situations. According to Kohler (1953), skill in performance must be achieved with the distorted vision before the new perceptual organization takes place. Although Kohler implies that performance in reversed or inverted visual fields may be equal to that found in normal vision, this conclusion is not clearly established by his or any other investigations. The effects of inversion may persist for a short time after the inverting optical systems are removed from the eyes.

This study is an attempt to develop new approaches to the problem of perceptual inversion. Instead of depending on subjective judgment of *S*s, scientific methods of motion analysis are used to assess performance with inverted vision. We have also attempted to achieve a relatively complete experimental design in observing important variables related to inversion. That is to say, a systematic theory has been used to plan the determination of the effects and interactions of task complexity, in relation to various conditions of visual and kinesthetic inversion. Finally we have eliminated secondary sources of optical error usually occurring in

[1] This research was supported by funds supplied by the National Science Foundation for a project on perception and human motion and in part by funds assigned by the Graduate School, University of Wisconsin.

[2] We are indebted to J. J. Gibson for supplying us with a mimeographed translation of the work of Kohler & Erisman.

inversion experiments due to the wearing of lenses or prisms. The *Ss* in this study viewed an inverted visual field through a fixed prism which produced no distortion of the visual field nor any movement of the visual field when the head is moved. On the basis of the general theory of human motions used in planning the study, certain predictions were made regarding the relative effects of kinesthetic inversion and visual inversion on the travel and manipulative movements of handwriting in handwriting tasks of different complexity.

METHOD

Apparatus.—In order to measure performance under the conditions of perceptual inversion, use was made of recently developed electronic methods of motion analysis (Smith & Bloom, 1956; Smith & Wehrkamp, 1951). Figure 1 illustrates the application of these methods to measurement of handwriting, as investigated in this study. The *S* is seated in front of an adjustable writing board over which a viewing aperature is mounted. He peers through this aperature in order to observe the movements of his hand in writing. To invert the visual field, a long right-angle (tank-type) prism is inserted in the aperature and *S* is required to view his performing hand through this fixed prism. A clear undistorted inverted visual field is obtained which does not move when *S* moves his head or eyes.

The *S* writes three separate letters or patterns—a series of dots (. . . .), a series of triangles (Δ Δ Δ Δ), and a series of "a's" (a a a a). The letters and patterns are written within a marked-off band of electrically conductive paper (Teledeltos[3] recording paper) which can be fed through the paper drive of the writing board. A metallic pencil containing electrographic lead is used.

As diagrammed in Fig. 2, a subthreshold current which passes through *S*'s body is keyed on and off by the contact of his pencil with the writing surface. Each time the pencil touches the paper it closes a relay circuit which operates a precision time clock and thus records the duration of this contact (manipulative) movement in writing. When the contact is broken by movement of the pencil to the next letter or pattern, the relay automatically activates a second clock which records the duration of this travel movement. As *S* writes each letter or pattern in a series, the apparatus automatically times separately and sums the two component movements of contact and travel. When the pencil comes to the edge of the paper, *S* touches a stop plate with his pencil and automatically stops the timing of the series of movements.

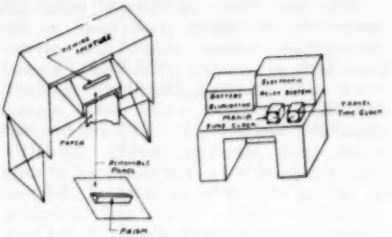

FIG. 1. The arrangement of the components of the electric handwriting analyzer. The writing apparatus, along with the removable panel containing the prism, is shown to the left. The relay and timing apparatus, shown to the right, are located on the opposite side of the experimental room.

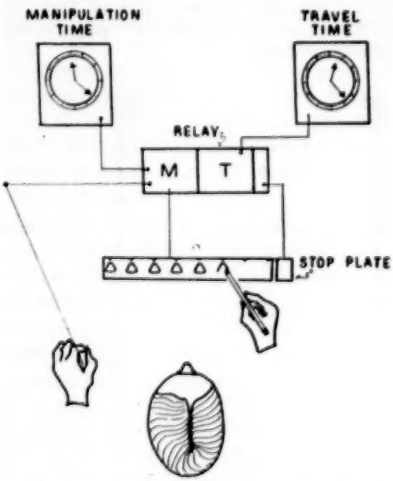

FIG. 2. The principle of the electronic handwriting analyzer.

[3] The Teledeltos Recording Paper used in this research is manufactured by the Western Union Telegraph Co.

Procedure.—Four conditions of visual and kinesthetic perception were used in the investigation—normal vision and normal kinesthetic orientation (NV-NKF), normal vision with inverted kinesthetic feedback (NV-IKF), inverted vision with normal kinesthetic feedback (IV-NKF), and inverted vision with inverted writing (IV-IKF). Four groups of 6 *Ss* each practiced 10 days in writing the three patterns of different complexity: dots, "a's," and triangles. University students were used as *Ss* and were rewarded for their performance by being assigned 10 points toward their final grade in elementary psychology.

On each day each of the patterns was written three times in three successive trials. Eight characters were written per trial. Practice effects were counterbalanced by having each *S* write the three symbols in a different sequence, such as dot, triangle, "a" or triangle, dot, "a." All six possible sequences were used for each of the four groups of six *Ss*.

In treating the data, the median contact time and median travel time of the three practice trials for each symbol on each day were used as the basis of all computations. Comparisons were then made between the performance of each group of *Ss* in the various experimental conditions on all of the symbols written or drawn. Analyses of variance of Day 1 performance and of Day 10 performance on each symbol and for both contact (manipulative) and travel time were made.

Results

Figures 3 and 4 describe the facts about adjustment to different conditions of perceptual inversion. Figure 3 gives the data for contact movements in the three task situations, and Fig. 4 the data for travel movements. The curves represent the duration of the component movements on successive days of practice. The parameter in each set of curves is the four conditions of perception used—NV-NKF, NV-IKF, IV-NKF, and IV-IKF. These curves show that the different conditions of perceptual inversion produced systematic variations in the duration of the contact and travel movements in writing. These variations are especially noteworthy in the writing of "a's" and triangles. The effects of inversion in tasks of different complexity are clearly seen. There are only limited effects of inversion on the writing of dots, an intermediate effect on writing "a's," and a maximum effect on writing Δ's.

These curves suggest that, initially, readjustment to inversion is somewhat faster for travel movements

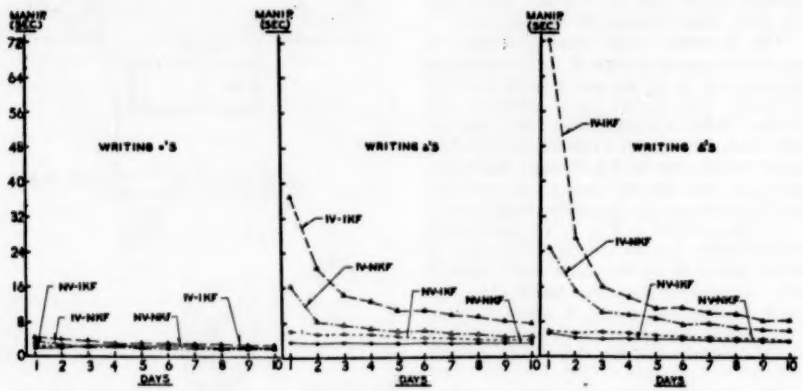

FIG. 3. Learning curves for manipulative (contact) movements in writing during readjustment to inverted perceptual fields.

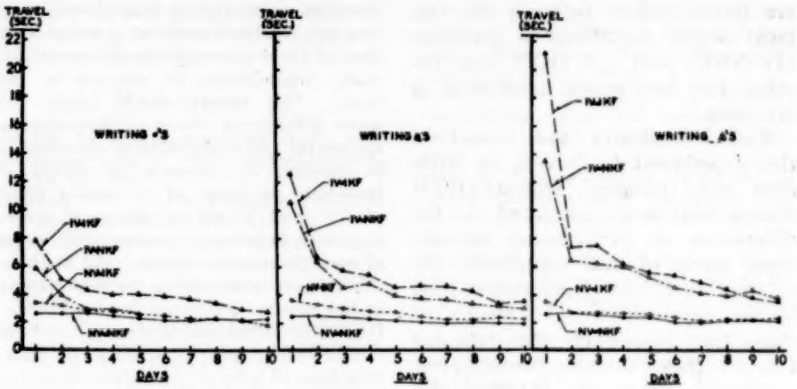

FIG. 4. Learning curves for travel movements in writing during readjustment to inverted perceptual fields.

than it is for manipulative movements. The curves for travel movements in the triangle task show relatively faster initial drop than those found for manipulative movements. Initially, the travel movements for the dotting task are more affected by the inversion than the contact movements.

There is an orderly variation of the effects of the four different conditions of inversion on both the contact and travel movements. The combined visual-kinesthetic inversion (IV-IKF) produces the most serious disturbance of movement with less serious effects appearing in the more limited conditions of inversion (NV-IKF and IV-NKF). Inverting the kinesthetic feedback of the writing motions (NV-IKF) has less effect on the movements than inverting the visual field (IV-NKF).

The significance of differences due to perceptual conditions has been assessed by means of analysis of variance[4] and Duncan (1955) Range tests made on the data of Day 10. Separate analyses were carried out for the contact and travel data. Results of these tests show that significant differences in performance exist especially between the condition of combined visual and kinesthetic inversion (IV-IKF) and the other conditions of perceptual inversion. For both contact and travel movements, significant differences in writing "a's" and triangles are found between the conditions of combined inversion (IV-IKF) and inverted writing (NV-IKF). Generally on this tenth day significant differences do not occur between the two conditions of visual inversion.

Similar range tests carried out for the data on Day 1 give results somewhat different from those for Day 10. For Day 1 manipulation times, differences are found mainly between the combined visual-kinesthetic inversion and the other three conditions. For Day 1 travel times, differences

[4] Six tables showing means of median manipulation and travel times by days (10), conditions (4), and task (3) and summaries of analyses of variance and trends are on file with the American Documentation Institute. Order Document No. 5846, remitting $1.25 for photocopies, or $1.25 for microfilms.

are found mainly between the two most severe conditions of inversion (IV-NKF) and (IV-IKF) and the other two less severe conditions of inversion.

Task complexity was varied in this experiment by having *Ss* write dots, "a's," triangles. Duncan (1955) Range tests were computed for the differences in performance at the three levels of task complexity for all four conditions of perception used in the experiment. The results of these tests, made on the data for Day 10, show, that for contact time significant differences between the dotting task and the other two tasks are found on both conditions of inversion of the visual field (IV-NKF and IV-IKF). A similar difference in travel time is found only for the visual inversion condition (IV-NKF). The absence of differences between "a" and triangle tasks at particular perceptual conditions does not mean that the writing of dots, triangles, or "a's" is unaffected by visual and kinesthetic inversion, but that the three tasks were not differentially affected in a significant way under a particular perceptual condition.

Discussion

Results of this study show that the initial and final adjustment to perceptual inversion depends upon the sensory mode and degree of such inversion, upon the component movements in the inverted performance, and upon the complexity of the task situation. With more complex tasks, both the initial and final effects of inversion are more severe. Adjustive recovery to visual inversion in more complex tasks is not complete after practice extending over 10 days.

The results of this experiment conform in detail to predictions based on the theory of space-structured movements in human motion (Smith, 1954). This theory interprets the adjustment to inverted vision and/or kinesthetic inversion as a matter of neuro-motor coordination of these systems in relation to primary organization of posture in the body. This theory would imply (*a*) some differential effects of inversion on the travel and manipulative movements in motion, (*b*) increase in effects of inversion on tasks of increasing complexity, and (*c*) greater effects of inversion with increase in amount and scope of such perceptual change. All of these effects were observed in the experiment in the manner predicted by the theory. It is to be noted that a theory based on reinforcement principles alone, or upon principles of general perceptual organization cannot be articulated to deal systematically with the variables of inversion specifically interpreted in this study.

The results found here show that the process of readjustment to perceptual inversion involves considerable specificity of the response in relation to movement components involved, in relation to level of task difficulty and in relation to the type and degree of inversion. This high degree of specificity of readjustment in the space organization and reorganization of motion under perceptual inversion is a clue to the specificity of psychomotor actions in general. We believe that the theory of space-structured movements in motion, which defined the plan and interpretation of this study, to be essential in giving an account of the specificity of perceptual motor performance under both normal conditions and under conditions of perceptual inversion.

Summary

Controlled experimental study involving assessment of the movements components in performance, and variation in the level of task difficulty and the type and degree of inversion, has been carried out to determine the effects of perceptual inversion on handwriting. Results show that readjustment to inversion is specific to all three aspects of performance investigated, as predicted by the theory of space-structured motion used to design and plan the study. In addition, the data and theory of the experi-

ment give a clue to the specificity which has been found to be characteristic of all psychomotor performance.

REFERENCES

BROWN, G. G. Perception of depth with disoriented vision. *Brit. J. Psychol.*, 1928, **19**, 117–146.

DUNCAN, D. B. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1–42.

EWERT, P. H. A study of the effect of inverted retinal stimulation upon spatially coordinated behavior. *Genet. Psychol. Monogr.*, 1930, **7**, 177–363.

EWERT, P. H. Factors in space localization during inverted vision. I. Interference. *Psychol. Rev.*, 1936, **43**, 522–546.

FOLEY, F. P., JR. An experimental investigation of the effect of prolonged inversion of the visual field in the rhesus monkey. *J. Genet. Psychol.*, 1940, **56**, 21–51.

HESS, E. Space perception in the chick. *Scient. Amer.*, 1956, **195** (1), 71–80.

KOHLER, I. Rehabilitation in perception. *Die Pyramide*, 1953, Heft 5, 6, and 7 (Austria). Special mimeographed translation, J. J. Gibson, Cornell Univer.

SMITH, K. U. General theory of human motions. Proc. 14th Int. Congr. of Psychol., Amsterdam, North-Holland Publishing Co., 1954, 184–185.

SMITH, K. U., & BLOOM, R. The electronic handwriting analyzer and motion study of writing. *J. appl. Psychol.*, 1956, **40**, 302–306.

SMITH, K. U., & WEHRKAMP, R. F. A universal motion analyzer applied to psychomotor performance. *Science*, 1951, **113**, 242–244.

STRATTON, G. M. Some repliminary experiments in vision without inversion of the retinal image. *Psychol. Rev.*, 1896, **3**, 611–617.

STRATTON, G. M. Vision without inversion of the retinal image. *Psychol. Rev.*, 1897, **4**, 341–360; 463–481.

STRATTON, G. M. The spatial harmony of touch and sight. *Mind*, N. S., 1899, **8**, 492–505.

WOOSTER, M. Certain factors in the development of a new spatial co-ordination. *Psychol. Monogr.*, 1923, **32**, No. 4 (Whole No. 146).

(Received June 12, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

THE SENSITIVITY OF THE EYE TO TWO KINDS OF CONTINUOUS TRANSFORMATION OF A SHADOW-PATTERN [1]

KAI VON FIEANDT
University of Helsinki
AND JAMES J. GIBSON
Cornell University

Many years ago, the French mathematician Poincaré (1913, p. 250 ff.) put a psychological question which has not yet been answered by psychologists: How does the eye distinguish between the change of *position* of a thing and the change of *state* of a thing? By "position," he meant the distance and orientation of an object in the environment, which can change when either the observer or the object moves. By "state," he meant the identity or rigidity of an object in the environment, which can change when the object grows or changes shape (even to the extent of becoming a different object). If vision presents to the eye only a projection or silhouette of the object, how does the eye discriminate between the *perspective* alterations of this projection and the alterations *arising from real changes of the object*? For example, how can the observer tell the difference between approach and growth, that is, between the object getting nearer and the object larger? How can he tell the difference between a turning motion of the object and a deformation of the object when, in both cases, its silhouette is compressed in one dimension? On the one hand, the silhouettes or visual sensations of objects are transformed when we move from one point of view to another, or when rigid objects themselves change position relative to the point of view. On the other hand, they are transformed when fluid or elastic objects change with respect to size or shape—a cloud or a surface of living tissue. But the changes of relative position are not confused in perception with the changes of structure or state. Men and animals seem to distinguish the rigid and permanent surfaces of the environment from the elastic or animate surfaces of the environment with considerable success. But why do both not look indifferently elastic since both may undergo transformation?

Poincaré, being a geometer, realized that the two kinds of transformations belong to different "groups." He thought that a man might *learn* to distinguish a change of position from a change of state by virtue of the fact that the first kind of transformation can always be reversed by a locomotor movement of the observer himself whereas the second kind of transformation cannot. That is, the observer can always move to a point of view which nullifies the first kind of transformation and re-establishes the original sensation and he can learn to associate his movement with the change of sensation. This is an interesting theory. However, a less speculative and more direct explanation is possible. It would be that an

[1] This experiment was performed at Cornell University where the first author was visiting on a grant from the U. S. Department of State. The first author designed the apparatus and carried out the experiment. The second author is responsible for the theory. We are indebted to Paul Olum of Cornell University for advice in the field of geometry.

eye, for reasons to be investigated, is sensitive to a change of form, and, moreover, to the *type* of a change of form. In this respect, an eye is to be contrasted with a camera. The eyes of animals and men are very good at detecting motions; perhaps they are just as good at discriminating *types* of motions. It might be that the one group of continuous transformations, the perspectives of rigid objects, constitutes one kind of stimulus for vision and that the other group, the "rubbery" transformations, constitutes another kind of stimulus for vision.

This hypothesis gains plausibility from the results of an experiment by Gibson and Gibson (1957) in which it was shown that a continuous perspective transformation (with polar projection) is a sufficient stimulus for an impression of rigid motion in space.[2] The transformation was carried by a pattern of shadows on the translucent screen of a point-source shadow projector (Gibson, 1957). The apparent surface seemed to rotate in the space behind the screen —that is, to change its position without change of its structure. But the experiment did not test the prediction that a continuous transformation of the other kind would give an impression of *non*-rigid motion—that is, a change of structure without change of position in space. The aim of the present experiment was to compare the perceptions induced by the two kinds of transformation sequences, everything else being equal so far as possible, and to determine whether corresponding impressions of rigid and of non-rigid motion would occur *spontaneously*.

[2] This research is allied to the stereokinetic phenomenon of Benussi and Musatti (1924) and to the kinetic depth effect of Wallach, and his collaborators (e.g., Wallach Weisz, & Adams, 1956).

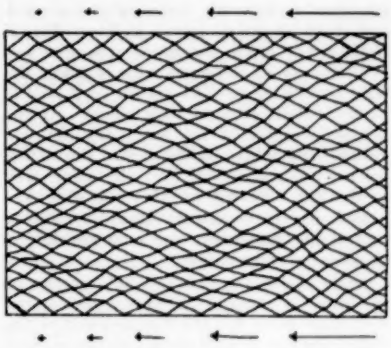

FIG. 1. The shadow-pattern. Note that the network is irregular and unfamiliar, and that a linear perspective cue is not available for the perception of slant. The change of pattern is indicated crudely by the arrows indicating displacements of elements. In the case of the perspective transformation, however, there is a vertical component of displacement as well as the horizontal component.

METHOD

Apparatus.—A point-source shadow projector is designed to produce an ongoing perspective transformation. It was necessary to modify the apparatus so that it would *also* produce an ongoing transformation of the "rubbery" kind, and so that it could be shifted from the first to the second kind of transformation without any realization by *O* that the apparatus has been changed.

For a decisive test, the two kinds of motion should be superficially similar but not mathematically similar. It was decided to fill the whole rectangular screen with a pattern (an irregular network of dark lines) which could be compressed from the right to the left of the window and then decompressed to restore the original pattern. There was to be always, in short, a gradient of leftward velocities of the elements of the pattern from a maximum on the right to zero on the left and then the same gradient of rightward velocities (Fig. 1). But one kind of compression of the image was to be the projection of a rigid surface which had been made to *turn*, whereas the other kind was to be the projection of an elastic surface which had been made to *contract*.

The shadow-caster behind the screen was therefore constructed in a special way. It consisted of a net of elastic thread stretched on a rectangular frame whose edges were

projected outside the frame of the window and were therefore never visible. The *E* could either rotate the frame around a vertical axis at its left-hand edge or allow the right-hand edge to slide inward on a pair of guides. Both motions were silent. At the top and bottom of the frame the net was attached by rubber bands so that the sliding motion was distributed over the whole area of the net. During rotation the slide was locked in its outermost position; during contraction the frame was kept parallel to the plane of the window.[3]

With this arrangement, a 20° rotation of the frame (toward the window) was equivalent to an 8-cm. inward displacement of the slide—equivalent with respect to the number of diamond-shaped elements of the shadows entering the window to the right. Stops were placed to limit the motions of the frame at these extremes. The *E* could thus, at will, produce either of two types of cyclical motion of the shadow pattern, operating the frame by hand at the rate of one cycle in 4 sec. The window was 38 × 40 cm. at a distance of 140 cm. from *O*'s eye, and the point-source was also 140 cm. behind the window. The frame was 39 × 47 cm. (when fully stretched) at a distance of 57 cm. behind the window (when parallel to the window). The *O* observed the window through a monocular aperture which hid the apparatus.

Procedure.—One way of comparing the perceptions aroused by two stimulus events is to determine whether the events are spontaneously named or identified with different words. This is not easy, however, if the ordinary language of the *O*s does not include accepted words for the different events. Another way is to determine whether the events are distinguished or discriminated by the consistent assignment of different words to different events, even though the words do not have accepted meaning. And a third way is to determine whether the transition from the presenting of one event to the presenting of the other arouses a reaction consistent with the direction of the transition. All three methods were used in the present experiment.

A naive *O* was brought into the dimly lighted experimental room, seated, and told:

"When you look through this tube you will see something. Please describe what you think it might be. When it begins to move, say what kind of movement it is. If the kind of movement changes, what new kind is it?"

The shadow of the net was motionless for a few seconds, and then cycles of motion were presented. Half the *O*s began with contraction and half with rotation. After three cycles (about 12 sec.) the other kind of motion was introduced without interruption for three cycles, and this alternation was continued until each kind of motion had been presented three times. During these six presentations *O* was freely reporting. This was the first part of the experiment, intended to find out if the verbal descriptions fell into two categories, and whether or not these could be taken to mean rigid and non-rigid motion. No suggestions were given at any time.

Nearly all *O*s (20 out of 22 tested) could be considered to have distinguished two classes of motion, although they had difficulty in finding words to describe what they saw. Using the terms spontaneously used by *O*, this group was then given the further instructions:

"From now on always use these names when you see the corresponding movements. Report whenever the movement changes from one to the other."

There followed 50 three-cycle presentations, 25 of contraction and 25 of rotation in a prearranged random order, half the *O*s beginning with one and half with the other. There were pauses after each trial so that the verbal identification could be recorded by *E*. This was the second part of the experiment.

Finally, a continuous series of motion-cycles was presented consisting of 47 transitions from contraction to rotation or the reverse, each transition being separated from the preceding by either three or four cycles of motion, these two possibilities being randomized. A record was kept of the changes reported or missed.

As an extra guarantee against the use of auditory cues for distinguishing the motions of the shadow-caster, a moderate masking noise was produced with a tape recorder placed near *O*.

Subjects.—Twenty *O*s completed the experiment, of which 18 were undergraduate students with little or no training in psychology and two were psychologists.

[3] The transformation of the shadow during lateral contraction of the frame was actually an approximation to the limiting case of a perspective transformation *with parallel projection instead of polar projection*. That is, it was similar to (but not wholly identical with) the motion which the shadow would have undergone if the frame had been rotated with the light source at a very great distance. This is a transformation which converts a square into a rectangle. With polar projection a square is converted into a trapezoid.

Results

The motionless shadow of the net was always seen as something in the frontal plane. The subsequent cycles of movement were described in a variety of terms whose meaning had to e interpreted. But the reports clearly fell into two types, and the change from one to the other movement of the shadow-caster was promptly noticed and reported.

The conclusion that the two kinds of transformation aroused two different perceptions is also warranted by the second part of the experiment. Out of 1000 judgments required of the *O*s, 979 or 98% were scored as correct, in that they consistently applied the terminology developed in the first six presentations.

The conclusion that the transitions were immediately responded to is proved by the third part of the experiment. Of 940 changes from rotation to compression or the reverse, all were responded to, and only one *O* on one occasion reported a change in the motion when none had occurred.

May we conclude that the *O*s spontaneously saw the same events which the *E*s saw in response to the two kinds of transformations? These events were, on the one hand, a rigid turning movement of an apparent surface in the space behind the window, and, on the other, an elastic or squeezing movement of an apparent surface, usually without change of depth but sometimes with a vague impression of it. The rigid rotation in depth could, with a shift of attitude, be seen as a compression, *but this was not the same kind of compression as the elastic movement.* The experience of compression obtained with the perspective transformation was, moreover, unstable.

The *O*s did not use the above terms to describe the events in question. They used familiar words like "wire fencing," "swinging toward me," "back and forth," "compression," "diagonal movement," "horizontal movement," and the like. The interpretation of such words is subject to error. Nevertheless the judgment of the authors is that *O*s in this experiment perceived essentially the same events. The perceptions differed in lacking the precision which can be achieved by using such terms as "rigid" and "elastic."

It is clear that, by two objective criteria, unsophisticated *O*s can distinguish between a continuous *perspective* transformation and a *nonperspective* transformation similar to it with respect to the gross pattern of displacements of elements. The visual system is sensitive to different kinds of change of projected form, at least in adults. This fact would explain the puzzle of Poincaré—why we do not confuse the rigid movements of surfaces in depth with the elastic movements of surfaces in the frontal plane. We cannot conclude with certainty that these *O*s did actually perceive a rigid surface in one case and a non-rigid surface in the other, for the extent to which the abstract property of rigidity was induced in our experimental situation is not clear. Nevertheless, the words used spontaneously by the *O*s suggest a tendency to do so.

Summary

Continuous perspective transformations of a rigid object and continuous nonperspective transformations of an elastic object were presented to *O*s under three conditions to determine whether discrimination of these would occur. All tests supported the inference that the discrimination is very efficient.

References

Gibson, J. J. Optical motions and transformations as stimuli for visual perception. *Psychol. Rev.*, 1957, **64**, 288–295.

Gibson, J. J., & Gibson, E. J. Continuous perspective transformations and the perception of rigid motion. *J. exp. Psychol.*, 1957, **54**, 129–138.

Musatti, C. L. Sui fenomeni stereocinetici. *Arch. Ital. Psicol.*, 1924, **3**, 105–120.

Poincaré, H. *The value of science* (Trans. by G. B. Halstead). New York: Science Press, 1913.

Wallach, H., Weisz, A., & Adams, P. A. Circles and derived figures in rotation. *Amer. J. Psychol.*, 1956, **69**, 48–59.

(Received June 27, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

REPLICATION REPORT: THE RELATIONSHIP OF MANIFEST ANXIETY AND ELECTRIC SHOCK TO EYELID CONDITIONING[1]

DONALD F. CALDWELL[2]
Vanderbilt University
AND
RUE L. CROMWELL
Peabody College

The present study partially replicates that of Spence, Farber, and Taylor (1954). Although apparatus was the same, certain changes were made in CR scoring and *S* assignment.

A response was scored as a CR during the 200 to 500 msec. interval after CS onset *only if it were preceded by a 200-msec. period of no response.* Rationale for this change was based on the fact that a voluntary blink initiated during the 200-msec. interval following CS onset is often carried over into the critical scoring range.

Sixty college students were divided into above- and below-median groups on MAS scores (Taylor, 1953). Within each group, pairs of *Ss* matched on MAS scores were assigned to "shock" and "no-shock" subgroups. The data was analyzed in two blocks of 40 trials each and with respect to the first 40 trials alone. This was done in two separate analyses. The first was a 2 × 2 × 2 design with High vs. Low Anxiety as a "between-effect" and Shock-No Shock and Trials 1–40 and 41–80 as "within-effects" (Lindquist, 1953, Type VII). A 2 × 2 table (Shock × Anxiety) (Lindquist Type I design) was analyzed with data of the first 40 trials only.

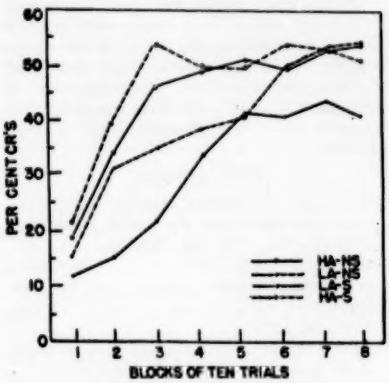

FIG. 1. Percentage of CR's in blocks of 10 trials.

Figure 1 presents the curves of acquisition. In the first analysis a significant difference was found only for trial blocks, indicating a greater number of CR's in the later trial block. No differences in learning could be attributed to the shock or anxiety dimensions. In the second analysis differences in learning were significantly related to differences in anxiety level only.

As in the original study, the present findings support the relationship between MAS and conditioning during the first 40 trials. The groups were not significantly different during the last 40 trials. Contrary to the original study, the present study finds no differences in conditioning as a function of shock. A further difference between the present and the original one seems to be the difference in level of asymptote. The *Ss* in the present study asymptoted approximately at or below a 50% CR rate; whereas, groups in the original study attained a 60–70% CR rate. Both findings can perhaps be accounted for by the difference in scoring procedures between the two studies. It is possible that the shock merely caused prolonged "voluntary" blinking which was scored positively in the Iowa study.

The authors are of the opinion that the use of a fixed interval (3 sec.) for the warning signal may also have given *Ss* an excellent cue to the onset of the CS (light) and thereby may have greatly increased the likelihood of the *Ss* voluntarily responding to the CS. In fact, many *Ss* reported upon completion of the session that they used the warning signal to avoid the puff.

If this criticism is cogent it may help account for the differences in the present study and the Iowa study in the amount of conditioning. The constant warning signal would have encouraged "guessing" in both studies and this guessing would be more sharply

[1] Acknowledgment is offered Leland Thune, Vanderbilt University, for his generous guidance.

[2] Previously at Peabody College.

penalized in the scoring procedure of the present study. Further substantiation for this notion comes from the fact that the frequency of CR's are more similar for the two studies during the earlier trials than during the later trials. In the earlier trials *S* would not have learned adequately the cue value of the warning signal. Therefore, voluntary or guessing responses as a function of the signal would be less expected here than during the later trials.

REFERENCES

LINDQUIST, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin, 1953.

SPENCE, K. W., FARBER, I. E., & TAYLOR, E. The relation of electric shock and anxiety to level of performance in eyelid conditioning. *J. exp. Psychol.*, 1954, **48**, 404–408.

TAYLOR, J. A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

(Received July 10, 1958)

Journal of Experimental Psychology
Vol. 57, No. 5, 1959

SUPPLEMENTARY REPORT: DISCRIMINATION LEARNING WITH PROBABILISTIC REINFORCEMENT SCHEDULES[1]

R. C. ATKINSON, W. H. BOGARTZ, AND R. N. TURNER

University of California, Los Angeles

This study deals with discrimination learning in a verbal conditioning situation and is an extension of research reported by Popper and Atkinson (1958). The study differs from theirs in that a new set of parameter values and a more rigorously controlled experimental procedure were employed. Two stimuli, T_1 and T_2, were used and two responses, A_1 and A_2, were available to *S*. Each trial began by the presentation, with equal probability, of either T_1 or T_2. On a T_1 trial, A_1 was correct with the probability π_1 and A_2 was correct with probability $1 - \pi_1$. For a T_2 trial, π_2 was identically defined.

Method.—Five groups were used. For all groups $\pi_1 = .9$. The groups differed with respect to the π_2 parameter which assumed the values of .9 (G-I), .7 (G-II), .5 (G-III), .3 (G-IV) and .1 (G-V). The *S*s were given 400 trials. However, on the first 40 trials all groups were given $\pi_1 = \pi_2 = .5$. The *S*s were 180 undergraduates, 36 per group. The experimental design was identical to that of Popper and Atkinson (1958) except that *S*s were run in subgroups of three, each *S* placed in a private booth after instructions. The apparatus, viewed from within *S*'s booth, consisted of two keys attached to the base of a panel, upon the panel were mounted four small lights. Two lights were in a column centered above the keys and served as the T_1 and T_2 stimuli. Each of the two remaining lights (the reinforcing signals) was mounted directly above one of the keys. On all trials the signal light (T_1 or T_2) was lighted for 1.5 sec.; the signal light went off simultaneously with the onset of a reinforcing light. The reinforcing light remained on for 1.8 sec., and was followed by an intertrial

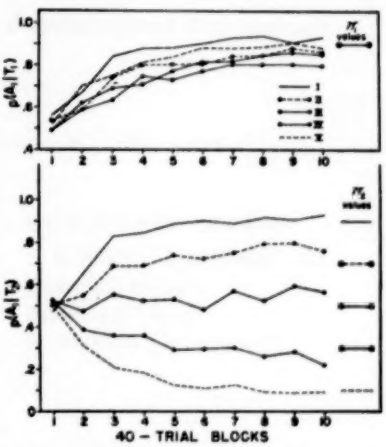

FIG. 1. Observed mean values of $p(A_1|T_1)$ and $p(A_1|T_2)$ in successive blocks of 40 trials. Each point is based on approximately 20 observations per *S*.

[1] This research was supported in part by a grant to the senior author from the National Science Foundation.

interval of 4 sec. Trials were run without interruption and S made his response while the signal light was on. Instructions were similar to those presented in Atkinson and Suppes (1958) except that they were modified to describe the T_1 and T_2 stimuli as in Popper and Atkinson (1958).

Results.—Figure 1 presents mean response curves over all trials of the experiment. In this figure the proportion of A_1's made on T_1 trials, $p(A_1|T_1)$, and the proportion of A_1's made on T_2 trials, $p(A_1|T_2)$, are plotted in successive 40-trial blocks. The corresponding π values are indicated on the far right. The asymptotic and pre-asymptotic characteristics of these curves are identical to those found by Popper and Atkinson (1958). One aspect of these results to be emphasized is the relation between the observed $p_\infty(A_1|T_1)$ and π_2. A convex function was found by Popper and Atkinson when π_2 varied from .85 to .15 (for a fixed $\pi_1 = .85$) and the same relation holds for this study. Specifically, if the proportions computed over the last 120 trials are used as estimates of $p_\infty(A_1|T_1)$, the obtained values are .930, .867, .808, .856, and .895 for Groups I to V, respectively. The convexity was found significant by evaluation of the quadratic component of the treatment sum of squares, $F = 18.2$.

The significance of these findings with regard to stochastic theories of discrimination learning is discussed by Atkinson (1958). In particular, the demonstrated convex relation between π_2 and $p_\infty(A_1|T_1)$ suggests that the Burke and Estes (1957) component model for discrimination learning is not applicable in this type of situation.

References

Atkinson, R. C. A Markov model for discrimination learning. *Psychometrika*, 1958, **23**, 309–322.

Atkinson, R. C., & Suppes, P. An analysis of a two-person game situation in terms of statistical learning theory. *J. exp. Psychol.*, 1958, **55**, 369–378.

Burke, C. J., & Estes, W. K. A component model for stimulus variables in discrimination learning. *Psychometrika*, 1957, **22**, 133–145.

Popper, J., & Atkinson, R. C. Discrimination learning in a verbal conditioning situation. *J. exp. Psychol.*, 1958, **56**, 21–25.

(Received August 22, 1958)

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:
UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:
J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts

Studies in Mathematical Learning Theory

Edited by
ROBERT R. BUSH and WILLIAM K. ESTES

This integrated collection of research papers considers the most recent developments in the analysis and application of mathematical learning theory. *Stanford Mathematical Studies in the Social Sciences, III.* $11.50

The Psychology of Affiliation

EXPERIMENTAL STUDIES OF THE SOURCES OF GREGARIOUSNESS
STANLEY SCHACHTER

In a series of masterly experiments, the author explores some of the major factors that determine our desires to be alone or with others. Aspects of affiliation and self-evaluation are examined under such conditions of stress as hunger, isolation, and fear of pain. *Stanford Studies in Psychology, I.* $3.75

The Gifted Group at Mid-Life

THIRTY-FIVE YEARS' FOLLOW-UP OF THE SUPERIOR CHILD
LEWIS M. TERMAN and MELITA H. ODEN

This is the fifth volume presenting results of the famous Stanford studies of gifted children. The subjects are now in their mid-forties, and the present study asks the question, "Have they in mid-life fulfilled the promise of their youth?" *Genetic Studies of Genius, Vol. V.* $4.50

STANFORD UNIVERSITY PRESS
Stanford, California

L. L. Thurstone

THE MEASUREMENT OF VALUES

With a preface by Thelma Gwinn Thurstone

L. L. Thurstone was one of America's outstanding psychologists, known for his studies of psychological measurements and for his application of precise measurement techniques to the psychological and social sciences.

Here are 27 of his most noted papers, arranged in five groups—learning, scaling, psychophysical measurement, attitude measurement, and multiple factor analysis. Each group is introduced with a brief discussion of the controversial issues current when the papers were published. The volume also contains a bibliography with a list of 95 Psychometric Laboratory Reports, a list of available microfilms and a partial list of psychological tests. $7.50

Through your bookseller

THE UNIVERSITY OF Chicago PRESS

Chicago 37. In Canada: *University of Toronto Press, Toronto 5, Ontario*

A *very considerable* proportion of all published articles on perceptual-motor skills and on perception-as-influenced-by-experience are given immediate publication in

PERCEPTUAL AND MOTOR SKILLS

If you are interested in these fields, you and your library should have current subscriptions to PMS ($10.00 per year, with a special rate of $5.00 to individual members of the APA). Prices of back volumes may be obtained on request from

PERCEPTUAL AND MOTOR SKILLS
PSYCHOLOGICAL REPORTS
Box 1441
Missoula, Montana